

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:47:57 ; Search time 17 seconds
(without alignments)
1721.378 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGPSEAEAPKAIPKTS.....RSFWIAIPWACPLGVAALC 562

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2770	92.7	1225	1 DAT1 HUMAN	Q9bco0 homo sapien
2	2110	70.6	614	1 DAT1 MOUSE	Q9c9b9 mus musculus
3	232	7.8	2039	1 PHF3 HUMAN	Q92576 homo sapien
4	203.5	6.8	656	1 CGBP HUMAN	Q9p0u4 homo sapien
5	194.5	6.5	660	1 CGBP MOUSE	Q9c9w7 mus musculus
6	193	6.5	488	1 CYL2 BOVIN	Q28092 bos taurus
7	189.5	6.3	705	1 TRDN RABIT	Q28620 oryctolagus
8	182.5	6.1	1010	1 HC58 MOUSE	Q8blg0 mus musculus
9	174.5	5.8	1012	1 HC58 HUMAN	Q9bvi0 homo sapien
10	174	5.8	700	1 TRDN CANFA	P82179 canis famil
11	174	5.8	704	1 NP14 RAT	P41177 rattus norv
12	172	5.8	5038	1 PCLO MOUSE	Q9qyx7 mus musculus
13	171.5	5.7	699	1 NP14 HUMAN	Q14978 homo sapien
14	169	5.7	728	1 TRDN HUMAN	Q13061 homo sapien
15	167	5.6	1101	1 PHF2 HUMAN	Q75151 homo sapien
16	164	5.5	1359	1 ATRX CAEL	Q9u7e0 caenorhabdi
17	163	5.5	1096	1 PHF2 MOUSE	Q9wcu0 mus musculus
18	161.5	5.4	1781	1 AKL2 HUMAN	Q02952 homo sapien
19	161	5.4	2492	1 ATRX HUMAN	P46100 homo sapien
20	161	5.4	2492	1 ATRX PANTR	Q7ygm4 pan troglod
21	161	5.4	2492	1 ATRX PONPY	Q7ygm3 pongo pygma
22	160.5	5.4	2004	1 MY33 HUMAN	Q92794 homo sapien
23	159.5	5.3	1087	1 NPH MOUSE	P19246 mus musculus
24	159	5.3	2476	1 ATRX MOUSE	Q61687 mus musculus
25	158	5.3	755	1 RREL HUMAN	Q92766 homo sapien
26	155	5.2	5085	1 PCLO RAT	Q9jks6 rattus norv
27	154.5	5.2	1861	1 MAP2 RAT	P15146 rattus norv
28	154	5.2	1191	1 CING MOUSE	P59242 mus musculus
29	153.5	5.1	6632	1 UN89 CAEL	Q01761 caenorhabdi
30	153	5.1	952	1 YK15 CAEL	P46012 caenorhabdi
31	153	5.1	1026	1 NPH HUMAN	P12036 homo sapien
32	151.5	5.1	848	1 NFM MOUSE	P08553 mus musculus
33	151.5	5.1	1338	1 ACIN MOUSE	Q9fjx8 mus musculus

ALIGNMENTS

RESULT 1

DAT1 HUMAN
ID DAT1 HUMAN STANDARD; PRT; 1225 AA.
AC Q9bco0; O15043; Q9BOW0; Q9BW03; Q9H4G6; Q9NTU8; Q9NCU8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Death associated transcription factor 1.
GN DAT1 OR KIAA0333.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Placenta;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugeno S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hoshiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saifu K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEBO human cDNA sequencing project.";
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.
MEDLINE=21638749; PubMed=11780052;
DELoukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,
Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
Huckle E., Hunt A.R., Hunt S.E., Jekesch K., Johnson C.M., Johnson D.,
Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
Lehvaeslalo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
Marsh V.L., Martin S.I., McConachie I.J., McClay K., McMurray A.A.,
Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
Skuce C.D., Smith M.I., Soderlund C., Steward C.A., Sulston J.E.,
Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
Tracey A.C., Tromam A.C., Vaudin M., Wall M., Wallis J.M.,
Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
Rogers J.;
The DNA sequence and comparative analysis of human chromosome 20.";
Nature 414:865-871 (2001).

DAT1_MOUSE STANDARD; PRT; 614 AA.
 AC Q89B9; Q89B2; Q9WV00;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Death associated transcription factor 1 (Death inducer-obliterating-1)
 DE (DIO-1).
 GN DAT1 OR DIO1 OR DIO1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RP TISSUE=Pre-B cell;
 RC MEDLINE=93324176; PubMed=10393935;
 RA Garcia-Domingo D., Leonardo E., Grandien A., Martinez P., Albar J.P.,
 RA Izpisua-Belmonte J.-C., Martinez-A.C.;
 RT "DIO-1 is a novel gene involved in onset of apoptosis in vitro, whose
 RT misexpression disrupts limb development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997 (1999).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Embryo, and Thymus;
 RC MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik J., Gough J.,
 RA Grimmond S., Guetlich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konoigaya A., Rurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ringwald M.,
 RA Sanderlin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 CC -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
 CC overexpressed.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
 CC after pro-apoptotic stimuli.
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
 CC -!- INDUCTION: Upregulated during apoptosis.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ238332; CAB48401.1; -;
 CC EMBL; AK042474; BAC31270.1; -;

DR EMBL; AK044919; BAC32141.1; -;
 DR MGD; MGI:1344352; Dicol.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0006915; P:apoptosis; IDA.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00628; PHD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS0016; ZF_PHD_2; 1.
 KW Apoptosis; Nuclear protein; Zinc-finger.
 FT DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 182 190 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT ZN FING 265 319 PHD-TYPE.
 FT CONFLICT 45 45 V -> A (IN REF. 1).
 FT CONFLICT 331 331 D -> N (IN REF. 1).
 FT CONFLICT 333 333 I -> V (IN REF. 1).
 SQ SEQUENCE 614 AA; 67409 MW; 15ACCSA4BAEF3539 CRC64;
 Query Match 70.6%; Score 2110; DB 1; Length 614;
 Best Local Similarity 75.2%; Pred. No. 4.9e-106;
 Matches 421; Conservative 30; Mismatches 95; Indels 14; Gaps 7;
 QY 1 MDKGDPSNEBAPKAIKPTSKFRKTWGRRTIARREGAGDAEADPLEPPPOOGLGS 60
 DB 1 MDKKGHLSNEBAPKAIKPTSKFRKTWGRRTIARREGAGDTEVDPSSQP--QHNLS 58
 QY 61 LRSGRQPKRTKRVQFLTIARRGRSRMPVSLDSGEPTSCPATDAETASEGSVASE 120
 DB 59 LRSGRQPKRTKRVVEFLTVARRGKNVPVSLDSSEPTSTVTDVETASEGSVASE 118
 QY 121 TRSGPOSASTAVKRPASSEKVKGGDDHDDTSDSDGLTKELQNRIRKRKEQTERP 180
 DB 119 IRSGFPVSDSLG-KEHPASSEKARGGEEDTSDSDGLTKELQNRIRKRKEQTERP 177
 QY 181 LKGISRLKRREREGPASTVGEASDTEVGLP-SKQEPENDQGVVSOAGKDDRESKLE 239
 DB 178 LRGSQNRLLAKRREREDSAST-GSVQIGSAEQRPCKQEPASQGVPSQSETDDIENQLE 236
 QY 240 GKAAQDIKDDEPDLGRPKPECEGVDPNALYICRQPHNNRPMICDRCEWPHGDCVGI 299
 DB 237 GRATQNTENPREAGKPKPECEVDPNALYICRQPHNNRPMICDRCEWPHGDCVGI 296
 QY 300 SGARGLLRNGEDYICPNTILQVDETHSETADQEAQKWPAGDADGTDCTSIGTIEOK 359
 DB 297 SGARGLLRNGEDYICPNTILQVDETHSETADQEAQKWPAGDADGTDCTSIGTIEOK 356
 QY 360 SSEDQIKIRIEKAANPSGKKLKIQQPVIEAPGASKICPGCGCHVAQPDVYCSNDCIL 419
 DB 357 SGEDQIKIRIEKAANPSGKKLKIQQPVIEAPGAPKICPGCGSVAPQPDVYCSNDCIL 416
 QY 420 KRAAATMKTLSGKEQKPKPKKKKKPKPSLPKCAQAGIKISSVHKRPAPKKEKTTV 479
 DB 417 KRAAATMKTLSGKEQKPKPKKKKKPKPSLPKCAQAGIKISSVHKRPAPKKEKTTV 476
 QY 480 KKAVVVPASSEALGKEACESSTPSWASDHNNVNAVPEKTAAP---SPSLLYKMYHLGV 536
 DB 477 KK-VMLASSETSGKEACESSTPSWASDHNNVNAVPEKPEKPTALSPTLISAKTYHPKA 535
 QY 537 GLLDPSRSFWIALPACPL 556
 DB 536 GFPGPSHLL-----GGLGL 550
 RESULT 3
 PHF3_HUMAN STANDARD; PRT; 2039 AA.
 ID PHF3_HUMAN
 AC Q92576; Q9NQ16; Q9UI45;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE PHD finger protein 3.
 OS Homo sapiens (Human).

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
PN	[1]
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX	MEDLINE=21845875; PubMed=11856865;
RA	Fischer U., Struss A.-K., Hemmer D., Michel A., Henn W.,
RA	Stedel W.-I., Meese E.;
RT	"PHF3 expression is frequently reduced in glioma.";
RL	Cytogenet. Cell Genet. 94:131-136(2001).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Bone marrow;
RX	MEDLINE=97191544; PubMed=9039502;
RA	Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y.,
RA	Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT	"Prediction of the coding sequences of unidentified human genes. VI.
RT	The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
RT	analysis of cDNA clones from cell line KG-1 and brain.";
RL	DNA Res. 3:321-329(1996).
RN	[3]
RP	SEQUENCE OF 83-2039 FROM N.A.
RA	Matthews L.;
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC	TISSUE SPECIFICITY: Ubiquitous. Expression is significantly
CC	reduced or lost in glioblastomas, glioblastoma cell lines,
CC	anaplastic astrocytomas, and astrocytomas.
CC	TISSUE SPECIFICITY: Contains 1 PHD-type zinc finger.
CC	TISSUE SPECIFICITY: This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/
CC	or send an email to license@isb-sib.ch).
DR	EMBL; AF091622; AAF21292.1;
DR	EMBL; D87685; BA13438.2; ALT_INIT.
DR	EMBL; AL050329; CAB93358.1;
DR	Genbank; HGNC:8921; PHF3.
DR	MIM; 607789;
DR	InterPro; IPR003618; TFS2 centre.
DR	InterPro; IPR001965; Znf_PHD.
DR	Pfam; PF00628; PHD; 1.
DR	SMART; SM00249; PHD; 1.
DR	SMART; SM00510; TFS2M; 1.
DR	PROSITE; PS01359; ZF_PHD_1; 1.
DR	PROSITE; PS50016; ZF_PHD_2; 1.
KW	Zinc-finger.
FT	ZN_FING 717 772 PHD-TYPE.
FT	DOMAIN 929 1029 TFS2M.
FT	DOMAIN 1757 1865 PRO-RICH.
SQ	SEQUENCE 2039 AA; 229480 MW; FDAFF00576005E9B CRC64;
Query Match 7.8%; Score 232; DB 1; Length 2039;	
Best Local Similarity 21.8%; Pred. No. 4.1e-05;	
Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;	
QY	92 SLED-SGPTSCPATDAETASGVSASER-SGQSASTAVKSPASSEKVK-GGDH 148
DB	595 SLSDSHAHFGCLPHEPHTQAGVSHSGKCHPQQQAPAKNTNSHVKELHPGVH 654
QY	149 DDTSDSDGLTXE-----LQNLRRKREQEPTERPLKGIQSLRKKRREGPARTVGE 204
DB	655 -----FKBEDKLLKPKPKLQPRRRSSFSLSDEPLFPDNIATIERE-----GS- 702
QY	205 ASDTVGVLPKQEPENDGVVSQAGKDRSKLGGKAAQIDKDEPGDLGPKPECEGY 264
DB	703 -----DHSSSPFSKY-----MWTSPKQC-GF 722
QY	265 DPNALYICICRPHNRFMTCCDRCEWFHGDGCVGISSEARGRLRNGEDYICPNTLIQV 324

RESULT 4

ID	CGBP_HUMAN	STANDARD;	PRT;	656 AA.
AC	Q9P004; Q96BC8; Q9P2V7;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	CpG binding protein (protein containing PHD finger and CXXC domain 1).			
GN	CGBP OR PCDL			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20153771; PubMed=10688657;			
RA	Shin Voo K., Carbone D.L., Jacobsen B.M., Flodin A., Skalniak D.G.;			
RT	"Cloning of a mammalian transcriptional activator that binds			
RT	unmethylated CpG motifs and shares a CXXC domain with DNA			
RT	methyltransferase, human trithorax, and methyl-CpG binding domain			
RT	protein 1.";			
RL	Mol. Cell. Biol. 20:2108-2121(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND DNA-BINDING.			
RX	MEDLINE=20261256; PubMed=10799292;			
RA	Fujino T., Hasegawa M., Shibata S., Kishimoto T., Inai S.-I.,			
RA	Takano T.;			
RT	"PCCL1, a novel DNA-binding protein with PHD finger and CXXC domain,			
RT	is regulated by proteolysis.";			
RL	Biochem. Biophys. Res. Commun. 271:305-310(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=21154917; PubMed=11230166;			
RA	Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glaesl S.,			
RA	Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,			
RA	Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,			
RA	Mewes H.-W., Ottenwaelter B., Obermaier B., Tampe J., Heubner D.,			
RA	Wambutt R., Korn B., Klein M., Poustka A.;			
RT	"Towards a catalog of human genes and proteins: sequencing and			
RT	analysis of 500 novel complete protein coding human cDNAs.";			
RL	Genome Res. 11:422-435(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Colon, and Skin;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haefl F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			

Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettanan M., Madan A.C., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schneringer A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[5]

DNA-BINDING DOMAIN.
MEDLINE=21576202; PubMed=11572867;
Lee J.H., Voo K.S., Skalniak D.G.;
"Identification and characterization of the DNA binding domain of CpG-binding protein.";
J. Biol. Chem. 276:44669-44676(2001).

-|- FUNCTION: Transcriptional activator that exhibits a unique DNA binding specificity for [AC]CpG[AC] unmethylated CpG motifs.
-|- SUBCELLULAR LOCATION: Nuclear.
-|- TISSUE SPECIFICITY: Ubiquitous.
-|- DOMAIN: The acidic domain carries the potential to activate transcription.
-|- PTM: May be regulated by proteolysis.
-|- SIMILARITY: Contains 1 PHD-type zinc finger.
-|- SIMILARITY: Contains 1 CXXC-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF149758; AAF37799.1; -
EMBL; AB031069; BAA96307.1; -
EMBL; AL1136862; CAB66796.1; -
EMBL; BC014940; AAH14940.1; -
EMBL; BC015733; AAH15733.1; -
GO; GO:0005634; C:nucleus; IDA.
GO; GO:0016563; F:transcriptional activator activity; IDA.
GO; GO:0045322; F:unmethylated CpG binding; IDA.
GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
InterPro; IPR002857; Znf_CXXC.
InterPro; IPR001965; Znf_PHD.
Pfam; PF00628; PHD; 1.
Pfam; PF02008; zf_CXXC; 1.
SMART; SM00249; PHD; 1.
PROSITE; PS01359; ZF_PHD_1; 1.
PROSITE; PS50016; ZF_PHD_2; 1.
Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc; Metal-binding; Coiled coil; Nuclear protein.
Zn FING 28 76
Zn FING 161 208
Zn FING 161 208
DOMAIN 256 317
DOMAIN 321 359
DOMAIN 422 474
DOMAIN 156 161
DOMAIN 162 221
DNA BIND 162 221
CONFLICT 117 117
CONFLICT 302 302
CONFLICT 340 340
SEQUENCE 656 AA; 57511 MW; 6D2376E849905A18 CRC64;
D -> N (IN REF. 1).
H -> N (IN REF. 1).
K -> KWME (IN REF. 4; AAH15733).

Query Match 6.8%; Score 203.5; DB 1; Length 656;
Best Local Similarity 24.0%; Pred. No. 0.00041;
Matches 82; Conservative 43; Mismatches 98; Indels 119; Gaps 20;

238 LEGKAAQIKDPERGDLGRPKPECEGYDPAALYCICQKQPHNRRFMICCDRCSEWFGDCV 297

[illegible]

```

or send an email to license@isb-sib.ch).
CC EMBL; AK010337; BAB26862.1; -.
DR MGD; MG1:1921572; Cdbp.
DR GO; GO:0016343; C:nuclear matrix; IDA.
DR InterPro; IPR002857; Znf_CXXC.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; ZF-CXXC; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
Metal-binding; Coiled coil; Nuclear protein.
FT ZN_FING 28 76 PHD-TYPE.
FT ZN_FING 165 212 CXXC-TYPE.
FT DOMAIN 260 321 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 325 363 ARG/LYS-RICH (BASIC).
FT DOMAIN 426 479 COILED COIL (POTENTIAL).
FT DOMAIN 156 161 POLY-GLN.
FT DNA_BIND 166 225 BY SIMILARITY.
SQ SEQUENCE 660 AA; 76166 MW; 9D92CB6DA82069EA CRC64;

Query Match 6.5%; Score 194.5; DB 1; Length 660;
Best Local Similarity 22.9%; Pred. No. 0.0012;
Matches 78; Conservative 46; Mismatches 104; Indels 113; Gaps 18;

QY 238 LEGKAAQDIKEEPPDGLGRPECEGYDPMALYICICROPHNNRPMICDRCZEFHGDVY 297
Db 1 MEG-----DGSLEPPDAG--DSSKSENGENAPIYICICRXPDPNCFMIGDCNENFHDGCI 55
QY 298 GISEARGLLERNGDYICPNC-----TILQVQ-----DETSETADQOEAQR 341
Db 56 RTERKAKAI-----REWCYRECKDKLEIRYRHKCRERDGERGSEPRDEGGGRK 111
QY 342 PGDADGTDCTSIGTIEQKSESDQIGRIEK--AANPSGKKKLIKIPQVIEAP----- 392
Db 112 P-----ASDPELQRRAGSGTGVGAMLRGASAPHSKSSP-----QPLVATPSQHHQOQ 159
QY 393 -----GASKCTGPGCCHVAQPDVSVCSNDCILKHAATMKFLSSGKQKPKPEKM 443
Db 160 QQQQQQIKRSARMC---GECEACRR-----TDC--GHCDFCRDMMKFGGPNKIKQKRL 209
QY 444 K-----MKPEKPS--LPKCGAQAQIKISSVHKRPAPKPKETTKVAVVY 485
Db 210 RQCLRARESYKVPSSLSPTVSEALPR-----PRRPPPTQOQ----- 248
QY 486 PARSEALGKEACSESTPSWASHNYNAVK--PEKTAAPSP 524
Db 249 PQOSOKLGRIDEDEGTVLSS-----SVVKEPPETATATPEP 282

RESULT 6
CYL2_BOVIN STANDARD; PRT; 488 AA.
AC Q28092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cylicin II (Multiple-band polypeptide II).
GN CYL2 OR CYL2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95255491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
heads: the identification and characterization of cylicin II.";
```

```

Exp. Cell Res. 218:174-182(1995).
[2]
RN PARTIAL SEQUENCE.
RP MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
the sperm head cytoskeleton."
J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: Possible architectural role during spermatogenesis. May
be involved in spermatid differentiation.
CC -!- SUBCELLULAR LOCATION: Calyx; sperm head cytoskeletal structure.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; Z46789; CNA86753.1; -.
DR PIR; I46014; I46014.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 487 49 X 3 AA REPEATS OF K-K-X.
FT DOMAIN 189 407 8 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 189 212 1.
FT REPEAT 213 239 2.
FT REPEAT 240 267 3.
FT REPEAT 268 295 4.
FT REPEAT 296 323 5.
FT REPEAT 324 351 6.
FT REPEAT 352 379 7.
FT REPEAT 380 407 8.
SQ SEQUENCE 488 AA; 53562 MW; 5F3107CB2D019C5A CRC64;

Query Match 6.5%; Score 193; DB 1; Length 488;
Best Local Similarity 20.3%; Pred. No. 0.0011;
Matches 117; Conservative 73; Mismatches 204; Indels 182; Gaps 19;

QY 18 PTKSEFFKTCGFRNTTIKREGAGDAEADPLEPPPPQQLGLSLRRSGROPKRTERVEQF 77
Db 19 PVSLSKKSN-----QQHTALVFPKPRPKR----- 46
QY 78 LTIARRGRSMPSVLSDESGEPTSCPATDAF-----TASEGSVESAS 119
Db 47 -----RSKPSLLQEN-----TSPKYDAEKLGRDRQPLMHRSLMRISERPSVYLAA 94
QY 120 ETR-----SGPOSASTAVKERPASSEKVKGGDDHDDTSDSDGLTLKELQRLRRKREOE 175
Db 95 RSRHPQKETPPSQBDAAQAAKPSKVKKSKEDKDKSDSEASISVSKPKPKLKAKEEK 154
QY 176 PTER-----PLKGTQS-----RLRKKRREBPATVSGSEASDT 208
Db 155 PDEKKDLKKRDKSKGKESATESDEKAGAEKGAEKDORKSGKKGKPTSDS--GSEKGA 213
QY 209 VEGVLPKQPEP-----NDQGVVSQAGKDDRESK-----LEGKAAQDIKEE 250
Db 214 KDKSKSKDKSKGKESATESSEGEKGDADKDDKDKGSKGKSKGKESATESSEGEKGDADK 273
QY 251 PGDLGRPK-----PECEGYDPMALYICICROPHNNRPMICDRCZEFHGDVCGISEARGR 305
Db 274 KKKGSKGKESATESSEGEKDAK-----DDKKGK 304
QY 306 LLENGEDYICPNTIILQVDETHSETADQOEAQRPGDADGTDCTSIGTIEQKSESDQ 365
Db 305 KSGKKGKE-----SATESGEKGDADKDDKDKGSKGK-----KESATESBG 347
QY 366 IKGRIEKAANPSGKKKLIKIPQVIEAPGASGICIGPCCHVAQPDVSVCSNDCILKHAAT 425
Db 348 EKGDAK--DDKKGK-----GSKK--GKESDKAEGDKGDAKD--DKDKKGS 391
QY 426 MKFTSSGKQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 485
```

```
Db 392 KIGKESATESEKSKODKAGKODPTKAGEKD---ESKOKKODAKKSKKKEKDEKK 448
Qy 486 PARSEALGKAAACESSTPSWASHDHYNAVPEKTA 521
Db 449 PGEARSEPKDSAKKODAKKODAKKODAKKDA 484

RESULT 7
TRDN RABIT
ID TRDN RABIT STANDARD; PRT; 705 AA.
AC Q28620; Q28636; Q28637; Q28643;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triadin.
DN TRDN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
OX [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RP STRAIN=New Zealand white; TISSUE=Skeletal muscle;
RC MEDLINE=91286104; PubMed=7685347;
RA Knudsen C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
RT "Primary structure and topological analysis of a skeletal muscle-
specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL J. Biol. Chem. 268:12646-12654 (1993).
RN [2]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RP TISSUE=Heart muscle;
RC MEDLINE=94298946; PubMed=8026576;
RA Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;
RT "Structural diversity of triadin in skeletal muscle and evidence of
its existence in heart.";
RL FEBS Lett. 348:17-20 (1994).
RN [3]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RP TISSUE=Heart muscle;
RC MEDLINE=96132942; PubMed=8550602;
RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT "Biochemical characterization and molecular cloning of cardiac
triadin.";
RL J. Biol. Chem. 271:458-465 (1996).
RN [4]
RN CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX MEDLINE=96066664; PubMed=7578102;
RA Fan H., Brandt N.R., Caswell A.H.;
RT "Disulfide bonds, N-glycosylation and transmembrane topology of
skeletal muscle triadin.";
RL Biochemistry 34:14902-14908 (1995).
CC - FUNCTION: May be involved in anchoring calsequestrin to the
junctional sarcoplasmic reticulum and allowing its functional
coupling with the ryanodine receptor.
CC - SUBUNIT: Homooligomer of variable subunit number; disulfide-
linked.
CC - SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
reticulum.
CC - ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=6;
Comment=Additional isoforms seem to exist;
Names=Skeletal 1; Synonyms=ST1;
IsoId=Q28820-1; Sequence=Displayed;
Names=Cardiac 1; Synonyms=CT1;
IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
Names=Cardiac 2; Synonyms=CT2;
IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
Names=Cardiac 3; Synonyms=CT3;
IsoId=Q28820-4; Sequence=VSP_004466;
Names=Skeletal 2; Synonyms=ST2;
IsoId=Q28820-5; Sequence=VSP_004462, VSP_004464,
VSP_004465;
CC
```

```
CC Name=Skeletal 3; Synonyms=ST3;
CC IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC - TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U31540; AAC48496.1; --
DR EMBL; L10065; AAC31488.1; --
DR EMBL; U31555; AAC48497.1; --
DR EMBL; U34201; AAC48498.1; --
DR PIR; A45990; A45990.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET 0
FT DOMAIN 1 46
FT TRANSMEM 47 67
FT DOMAIN 68 705
FT CARBOHYD 74 74
FT CARBOHYD 624 624
FT VARSPLIC 264 285
FT VARSPLIC 264 307
FT VARSPLIC 286 705
FT VARSPLIC 308 705
FT VARSPLIC 415 415
FT VARSPLIC 416 424
FT VARSPLIC 584 584
FT VARSPLIC 585 591
FT VARSPLIC 645 705
FT
FT QUERY MATCH 6.3%; Score 189.5; DB 1; Length 705;
FT BEST LOCAL SIMILARITY 22.9%; Pred. No. 0.0024;
FT MATCHES 99; Conservative 62; Mismatches 177; Indels 95; Gaps 16;
Qy 145 GDDHDTSDSDSDELTLKE--LQNLRLRRRBEQPTERPLKGIOSRL---RKKEREKGP 199
Db 117 GDEDDGDEDTAKGEIEBPPLKRDIIHEKIEKQEKPERKPTKVVHKEKEKKEKVKVE 176
Qy 200 TVGSEASDVTVEGVLPSKQSPENDQGVVSOAGKDDRESKLEGAQAQDIKDEPGDLGR--- 256
Db 177 KEKPEKKATHKEKLEKKEKPEP-KTVTKEKKARTKEKIEKTEKKEKVKQKVKQTVA 235
Qy 257 -----PKP-BCEGYDPNALYCICQPHNNRPMI CDRCEWTFGDCVGISEARGRL 307
Db 236 KAKEVQKTPKPEKESKETAA---VSKQEQKQYAFCRYMIDIFVHGD-----L 281
Qy 308 ERNGEDYICPNCITLQVDETHS-ETADQEAARPGDADGTDCTGCTGIGTIEQKSSDOGI 366
Db 282 KPCQSPALPPSPTEQASRPTPALPTPEKEGKKKAB-----KVTETTKKAEKEDA 335
Qy 367 KGRIRKAAKNSGKKLKIIFQPVIEAPGASKICIGPGCHVAQPDPSVYCSNDCILKHAATM 426
```


childhood medulloblastoma cDNA libraries.";
Int. J. Cancer 106:244-251(2003).
-!- FUNCTION: Possible transcription factor.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).
-!- ALTERNATIVE PRODUCTS:
Event-Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q9BVI0-1; Sequence=Displayed;
Name=2;
IsoId=Q9BVI0-2; Sequence=VSP_007760, VSP_007761;
Note=Ref.1 isoform 2 is in conflict in position(s):
237-239:KTR>KXTS;
-!- TISSUE SPECIFICITY: Expressed in heart, kidney, liver, lung,
pancreas, placenta, spleen and testis. Not expressed in brain,
skeletal muscle, colon, ovary, prostate, small intestine and
thymus. Expressed in colon and ovary cancer cell line, whereas it
is not expressed in the respective normal tissues.
-!- DISEASE: Antibodies against C20orf104 are present in sera from
patients with hepatocellular carcinoma, glioblastoma and childhood
medulloblastoma who developed different autoantibodies. May serve
as diagnostic marker in glioblastoma.
-!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
-!- SIMILARITY: Contains 1 PHD-type zinc finger.
-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 260.
-!- CAUTION: Ref.2 (AAK19748) sequence differs from that shown due to
frameshifts in positions 635, 653 and 655.
-!- CAUTION: Ref.4 sequence differs from that shown due to a chimeric
cDNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF220416; AAF34184.1; ALT_FRAME.
EMBL; AY027523; AAK13046.1; -.
EMBL; AF348207; AAK19748.1; ALT_FRAME.
EMBL; AL078461; CAC09384.1; -.
EMBL; AL109965; CAC09389.1; -.
EMBL; AF258787; AAG49888.1; -.
EMBL; BC001198; AAH01198.1; ALT_SEQ.
EMBL; BC048210; AAH48210.1; -.
Genew; HGNC:16098; C20orf104.
InterPro; IPR002999; Tudor.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR001965; Znf_PHD.
Pfam; PF00628; PHD; 1.
Pfam; PF00096; Zf-C2H2; 1.
SMART; SM00249; PHD; 1.
SMART; SM00333; TUDOR; 2.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.
PROSITE; PS01359; ZF_PHD_1; 1.
PROSITE; PS00016; ZF_PHD_2; FALSE NEG.
Transcription regulation; Nuclear protein; DNA-binding; Metal-binding;
Zinc; Zinc-finger; Alternative splicing.
FT DNA BIND 257 269 A.T HOOK.
FT ZN FING 452 477 C2H2-TYPE.
FT ZN FING 654 700 PHD-TYPE.
FT DOMAIN 162 221 LYS-RICH.
FT DOMAIN 523 552 LYS-RICH.
FT VARSPLIC 237 273 VDKPNDIVKSPQENLRPKRGPPSIAPTVDS ->
KTRQPFHSSVCCGFKLSFNATNIGTEKENIKEM (in
isoform 2).
FT FT /FTId=VSP_007760.
FT VARSPLIC 274 1012 Missing (in isoform 2).
FT FT /FTId=VSP_007761.
FT CONFLICT 221 222 KE -> QG (IN REF. 5).

FT CONFLICT 226 226 E -> G (IN REF. 5).
FT CONFLICT 516 516 S -> F (IN REF. 5).
SQ SEQUENCE 1012 AA; 115385 MW; 1CDBADDC23D007503 CRC64;
Query Match 5.8%; Score 174.5; DB 1; Length 1012;
Best local Similarity 19.3%; Pred. No. 0.023;
Matches 156; Conservative 72; Mismatches 269; Indels 285; Gaps 32;
QY 3 DKGDPNREAPKAKPTSKFRITWGRFRRTTAKREGAGDAEADPLEP-----PPPQQQL 57
DB 204 EKGVSEKSLPKNEKEDKENISEN-----DREVSQDAQVDK-KPENDIVKSPQENL 253
QY 58 GLSLRSGROPK-----RTVEQFLITARR----- 84
DB 254 REPRKRGPPSIAPTVDSNSQTLOPITLERRKISKGCYVLPKRPRLDKNSSQBSKSK 313
QY 85 -----GRSMPVSLDSGE-----PTSCPATDAETASEGVSESA 118
DB 314 NYSENTDKLSRRSRSLSTNGTHEILDPLVVDLVDTPDLOUTLSSTKESEGQKLSA 373
QY 119 SET-----RSG-----POSASTAVKERPASS-----EKYGGDDHD 149
DB 374 LEAQVSSALTCFSGDGAAGLELNCPSMGENTMTKTEFTSLVLEQLEISTVEVTNFK 433
QY 150 DTSDDSDGUTLKEQLNRLRK-----REQEPTEPLK-G 183
DB 434 KTDGSSNAPAVDLDHFKCKVVDCLKPRKAKLLHYHMKYTHGMEKSLSESPGKRH 493
QY 184 IQSR-----LRKGRREGP 197
DB 494 VQTRGPSADKPSQETLTRKRVSSASPTTNDKRNKEKFKFVVRVKKKKKKKKTKP 553
QY 198 AETVGSASDTVGVLPSK-----QBPENDQGVWSQAGKDD- 233
DB 554 ECPSEISITSEPPFPKAFATRCGSSHKPGVHMSPOLHGPESHGKVKVXALBEDNL 613
QY 234 RESKLEKAAQDIKDEPPGDLG---RPKPCEG---YDPNALYICIQPHNRRFMICDR 287
DB 614 SSSSSSFLMSD--DEYGDQVDVTNPDDELGDGDYDFEVVRCICEVQENDFMIOCEE 671
QY 288 CEWFQDCVCISEARGRLERN-GDYICPNTILQVQDETHSETADQAEAKWRPDAD 346
DB 672 CQCMQHGVCNG-----LLENVPEKITCYVC-----QDPGQRPFGKY 709
QY 347 GTD-----CTSIGTIEOKSSEDQIK-----GRIEKAAN--PSGKKKLIKIPQV 388
DB 710 WYDKWLSRGHMGFLAFLEENYSHONAKKIVATHQLIGDVQVRIEVLHGLQKMSILQS- 768
QY 389 IEAPGASKICPGCCHVAQPDVYCS---NDCILKHAATWKLSCGKCKEKKMKM 445
DB 769 RHPDPLPWCOPWKQHSGEGRSHFRNIPVTDTSKEEAPSYRIL-NGAVEKPREL- 822
QY 446 KPEKPSLPKCAQAGIKISSVHKRPAPKEKETTVYKAVVVPVPAVSEALG-----KEAAE 499
DB 823 -----ALPLPRSVESYITSEHCYQKPRAYPAVBQKLWVETRGSALDDAVNPLHENGDD 877
QY 500 SSTP--SWASHNYNA-----VKP-----EKTAAPSLLYKCMVHLGVGLDPSRS 544
DB 878 SLSPGLGWFLDQDRSGKSDPKFGSPKVKYVSKALPSEAPARKLLDRGSEGLLS-SQH 936
QY 545 FW 546
DB 937 QW 938
RESULT 10
TRDN CANPA STANDARD; PRT; 700 AA.
ID TRDN CANPA
AC P82179;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triadin.

```

QY      219 PNDQGVVSQAQKDRBSKLEGAQAQDIKDREPGDLGRPKP-----ECF 262
DB      192 PE-----TKTWAKERKKARTESIKKEVKG-----GQEKVKPTAAKVKEVQKTTPPAKEKE 243
QY      263 GYDPNALYCICRPHNNRFMICCDRCBEPFHGD--VGISEARGLLERNGEDYICPNCTI 321
DB      244 GKETA---VAHQEQDQVAFCRYMIDMFVHGLRGQSPALPPPLPTVOASRPPTASPT 300
QY      322 LQVQETHSETADQSEA-----KWPCDADGTDCSTSGTI- 356
DB      301 LEQEIEEEKKAEKAVTGETKKCKEDVKKSKDKDTAIDVEKKCPKAPETKQGITLKVVA 360
QY      357 -----EOKSSSDQIGIKRIKAANPSGGKKLKIPOPVLEAPGASCKICGGCHVAQPDS 410
DB      361 QAAAANKDEKESKTKTFVEZ--HPKGGKQEKKEKYVEPASKK-----EHSAPSEK 412
QY      411 VYCSNDCILKH---AAATKFSSGQEKPKPEKKMKMPKPESLPKCGAQAGIKISSVH 467
DB      413 QVKATERAKEETSAASTKAVGPKKEERTTKTVGEIRKESG-----KTSTAS 462
QY      468 KRPAPR-KETT VKA--VVVPARSEA-LGKEAACESSTFVASDENYNNAVKPEATAA 521
DB      463 KDEPTEIMKDEKMPKADKEVKPRPPQSQVQKBEKSESQVKKAKPEQ-DIAKPEKTIVS 519

RESULT 11
NP14_RAT
ID NP14_RAT STANDARD; PERT; 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE phosphoprotein 1).
DE NCBI_TaxId=10116;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX [1]_TaxId=10116;
RN Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
RN phosphoprotein 1).
RP SEQUENCE FROM N.A., AND SEQUENCE OF 232-309 AND 563-601.
RC TISSUE=Liver;
RC MEDLINE=922323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL Cell 70:127-138(1992).
RN [2]
RP INTERACTION WITH NOP5 AND FIBRILLARIN.
RX MEDLINE=20143579; PubMed=10679015;
RA Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT "Conserved composition of mammalian box H/ACA and box C/D small
RT nucleolar ribonucleoprotein particles and their interaction with the
RT common factor Nopp140."
RL Mol. Biol. Cell 11:567-577(2000).
CC -! FUNCTION: Related to nucleogenesis, may play a role in the
CC maintenance of the fundamental structure of the fibrillar center
CC and dense fibrillar component in the nucleolus. It has intrinsic
CC GTPase and ATPase activities. May play an important role in
CC transcription catalyzed by RNA polymerase I (By similarity).
CC -! SUBUNIT: Interacts with DKC1/NAP57, NOP5/NAP65 and fibrillarin.
CC -! SUBCELLULAR LOCATION: Shuttles on curvilinear tracks between
CC nucleolus and cytoplasm. These tracks extend from the dense
CC fibrillar component of the nucleolus across the nucleoplasm to
CC a limited number of nuclear pore complexes.
CC -! PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
CC cycles on CK2 and PKC sites. Nopp140 is one of the mostly
CC phosphorylated proteins in the cell.
CC -! SIMILARITY: Contains 1 Lish domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```


RL Biochem. Biophys. Res. Commun. 230:370-375 (1997).
RN CHARACTERIZATION.
RX MEDLINE=20036810; PubMed=10567578;
RA Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;
RT "Human Nopp140, which interacts with RNA polymerase I: implications
for rRNA gene transcription and nucleolar structural organization.";
RI Mol. Cell. Biol. 19:8536-8546 (1999).
CC -!- FUNCTION: Related to nucleogenesis, may play a role in the
maintenance of the fundamental structure of the fibrillar center
and dense fibrillar component in the nucleolus. It has intrinsic
GTPase and ATPase activities. May play an important role in
transcription catalyzed by RNA polymerase I.
CC -!- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPA194)
and with casein kinase-II.
CC -!- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the
cytoplasm. At telophase it begins to assemble into granular-like
pre-nucleolar bodies which are subsequently relocated to nucleoli
at the early G1-phase.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha;
CC IsoId=Q14978-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q14978-2; Sequence=VSP 004338;
CC -!- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
cycles on CK2 and PKC sites. There is evidence suggesting that
CDK2 kinase phosphorylates p130 at the N-phase.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z34289; CAA84063.1; -;
CC EMBL; D21262; BAA04803.1; -;
CC PIR; I38073; I38073
CC Genew; HGNC:15608; NOLC1.
CC GK; Q14978; -;
CC MIM; 602394; -;
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005730; C:nucleolus; TAS.
CC GO; GO:0007049; P:cell cycle; TAS.
CC GO; GO:0007067; P:mitosis; TAS.
CC GO; GO:0006364; P:rRNA processing; TAS.
CC InterPro; IPR006594; Lish.
CC InterPro; IPR007718; SRP40_C.
CC Pfam; PF05022; SRP40_C; 1.
CC SMART; SM00667; Lish; 1.
CC PROSITE; PS50896; Lish; 1.
CC Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW Alternative splicing; 42
FT DOMAIN 10 42 LISH.
FT DOMAIN 84 566 11 X 12 AA APPROXIMATE REPEATS OF AN
FT REPEAT 84 95 ACIDIC SERINE CLUSTER.
FT REPEAT 125 136 ACIDIC SERINE CLUSTER 1.
FT REPEAT 167 178 ACIDIC SERINE CLUSTER 2.
FT REPEAT 221 232 ACIDIC SERINE CLUSTER 3.
FT REPEAT 264 275 ACIDIC SERINE CLUSTER 4.
FT REPEAT 325 336 ACIDIC SERINE CLUSTER 5.
FT REPEAT 363 375 ACIDIC SERINE CLUSTER 6.
FT REPEAT 425 436 ACIDIC SERINE CLUSTER 7.
FT REPEAT 470 481 ACIDIC SERINE CLUSTER 8.
FT REPEAT 519 529 ACIDIC SERINE CLUSTER 9.
FT REPEAT 555 566 ACIDIC SERINE CLUSTER 10.
FT DOMAIN 68 82 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 204 382 INTERACTS WITH RPA194.
FT DOMAIN 384 587 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT

FT DOMAIN 601 617 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 563 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT VAR_SEQ 241 K -> KWTITSVRAE (in isoform Beta).
FT /FTID=VSP 004338.
FT CONFLICT 3 3 D -> A (IN REF. 2).
FT CONFLICT 133 133 R -> S (IN REF. 2).
FT CONFLICT 291 292 YA -> SV (IN REF. 2).
FT CONFLICT 456 456 S -> P (IN REF. 2).
SQ SEQUENCE 699 AA; 73720 MW; DFDAD94EDF659FB CRC64;
Query Match 5.7%; Score 171.5; DB 1; Length 699;
Best local Similarity 19.5%; Pred. No. 0.022;
Matches 108; Conservative 84; Mismatches 218; Indels 145; Gaps 19;
QY 2 DDKGDSNEAPKAIKETSKEPKTWGRTTIAKREGAGDAEAD-PLRPPPPPOOQLGLS 60
DB 136 DDEEDKQKQVQGVKPPQAKAARAP-----PKAKSSDSDSDSSEDEPPNQ----- 183
QY 61 LRSSGQPKTERVEQFLTIARRRRRRMSPVLSDEGPTSCPATDAETASRGSVASE 120
DB 184 -----KPKIT-----PVTVAQTAKPPKAPARAAPKIANGKAASSSS 219
QY 121 TRSGPQASTAVKRPASSEK-----VKGDDHDDTSDSDSLTLKEIQNR 167
DB 220 SSSSSSSSDSDEEKAATAATPKTVKQVAVKAPVKAATPTTKSSSDSDSDEEBOQ 279
QY 168 LRKREQEP-----TERPLKGIOSRLKQRRGEAEFTVGEASDITVEGVLPKQEP 219
DB 280 KPMNKPGPVSAPPSPAPPKSLGTQPPKAVEKQPVSESSDSDSDSSEEEKKP 339
QY 220 ENDGVVSGKDDRESKLGKAAQDIKEPDGLGRPKCEGYDPNAL----- 269
DB 340 FT-KAVVSKATTFPPAKAAESSSDSDSDSDEDDRAPSPAGTTKNSSNKPATTKSP 398
QY 270 ----YCIQCPHNNRFWICCDRCBEPHGCVGISREARGLERNGEDYICPNCITILQVQ 325
DB 399 AVFPAAPKQP-----VGGGQ---KLLTKAD-----SSS 425
QY 326 DETHSETADQOEAK-----WRPDADGDTCTSI-----GTIEQKSEDOGIKRIEKA 373
DB 426 SEESSSSSEBEKTKKVVATTKP-KATAKAALSFPKQAPQSGRSDSDSDSSEBEHK 484
QY 374 ANPSGKKLKIQFQVIEAPGASKICIGPCCHVAQDPVYCSNDCILKHAATMKFLSGSK 433
DB 485 TSKSAVKK-----KQKVAGGAP-----SKPASA-----KKGKAESSNSSSD 523
QY 434 EQKPKPKKMKOK-PEKPSLPKCGAAGIKISSVHKRPAPKPKETTVKKAIVVPARSEAL 492
DB 524 DSSEEEKLGKSGSPRQAPKANGTSALTQAQKAAKNSEEEBEKKKAIVVVSXGSL 583
QY 493 GK-----EAACESSTP 503
DB 584 KKKQNEAAKAEATP 598
RESULT 14
TRDN HUMAN STANDARD; PRT; 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Syre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,

RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RL and its localisation to chromosome 6q22-6q23.";
CC Eur. J. Biochem. 233:258-265 (1995).
CC
CC -!- FUNCTION: May be involved in anchoring calsequestrin to the
CC junctional sarcoplasmic reticulum and allowing its functional
CC coupling with the ryanodine receptor (By similarity).
CC
CC -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-linked
CC (By similarity).
CC
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18985; AAB5315.1; -.
CC PIR; S68191; S68191.
CC Genew; HGNC:12261; TRDN.
CC MTM; 603283; -.
CC GO; GO:0016021; C:integral to membrane; TAS.
CC GO; GO:0006936; P:muscle contraction; TAS.
CC Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
CC INIT MET 0 0 BY SIMILARITY.
CC FT DOMAIN 1 46 CYTOPLASMIC.
CC FT TRANSXEM 47 67 POTENTIAL.
CC FT DOMAIN 68 728 LUMENAL.
CC FT DISULFID 269 269 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 690 690 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 74 74 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 646 646 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 728 AA; 81423 MW; C1C53BB1B2A0815 CRC64;

Query Match 5.7%; Score 169; DB 1; Length 728;
Best Local Similarity 19.3%; Pred. No. 0.032;
Matches 101; Conservative 83; Mismatches 168; Indels 172; Gaps 22;
QY 90 PVSLESGEPTSPATDAETASGVESA-----SET-----RSQPSASTAVKE 134
DB 292 PLTEQAASRT--PASPALKEKEKKAKKAVTSTKXKKEKEDIKKKSKETAIDVKK 349
QY 135 RPASSEKVKGS-----DDHDTSDSDSGTLKLNRLRRKEQ--EPTERP 180
DB 350 EPKASSETKGTGKIAAQAANKDEKEDSKTKKPAVEQPKKKKKKKVEPAKSP 409
QY 181 LK--GIOSRLRKKRRERGPATVGEASDTVGVLPKQBPENDQGVVSOAGKDDRESKL 238
DB 410 KKEHSVPDDQVAKTERAKKEIGAVS-----KKAAPGKBEKTKTVEQIRKEK----- 461
QY 239 EGKAAQIKDEE-----PGDLGRPKPCBGYDPNALYCIQPHNNRPMICDRCR 289
DB 462 SGKTSILKDKPTIKGKEKVPASLAKKEPETK----- 494
QY 290 EWFHGDGCVIGSEARGILLRNGEDIYICPNTLIQVODETHSETAQOQAKWRFGDADGTD 349
DB 495 -----XDEKMSKAG--KEVKPKPPQLQK- 516
QY 350 CTGIGTTEQKSSPDQKIGRIKKAANPSGKKKLIKIPFVIEAPGASKICIGFGCHVAQPD 409
DB 517 -----KEEKPEPQ-----IKKEAPKPAISQVQIHQDVKP--EKTVSHG-----KPE 557
QY 410 SVYCSNDILKHAATMKFLSSGKQKPKPKKKK-----MKPEKPSLPKCG-----A 457
DB 558 -----EKVLQ-----VKAETKAKPKTKAEEHREPPPSIKTDKPKTKGTSEVT 607
QY 458 QAGIKSSVHKRPAPKEKTKVKAUVVPARSEALGKEAACESSTPSWASD--HNTYNAVX 515
DB 608 ESKGKTKTEISKESKEKADMKHLREKVKVSTRKESLQHLNVTKAKPARVSKVEDVPASK 667

QY 516 PKK-----TAPSPSLLYKCMHGLGVLLDPSRSFWIAIPW 551
DB 668 KAREGTEDVSPTKQKSPISFFQCVY-----LDGNGYGFQFPF 705
RESULT 15
ID PHF2_HUMAN STANDARD; PRT; 1101 AA.
AC 075151; Q8N3K2; Q9Y6N4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PHD finger protein 2 (GRCS).
GN PHF2 OR KIAA0682.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99160472; PubMed:10051327;
RA Hasenpusch-Theill K., Chadwick B.P., Theill T., Heath S.K.,
RA Wilkinson D.G., Frischauf A.M.;
RT "PHF2, a novel PHD finger gene located on human chromosome 9q22.";
RN Mamm. Genome 10:294-298(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:98403880; PubMed:9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RN DNA Res. 5:169-176(1998).
RN [3]
RP SEQUENCE OF 2-1101 FROM N.A.
RC TISSUE=Amalgama;
RA Koehler K., Beyer A., Meves H.-W., Weill B., Wiemann S.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 JMC domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF043725; AAD21791.1; -.
CC EMBL; AB014562; BAA31637.2; ALT_INIT.
CC EMBL; AL834263; CAD38938.1; -.
CC Genew; HGNC:18920; PHF2.
CC MIM; 604351; -.
CC InterPro; IPR003347; TF_JmJC.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF02373; JmJC; 1.
CC Pfam; PF00628; PHD; 1.
CC SMART; SM00558; JmJC; 1.
CC SMART; SM00249; PHD; 1.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS50016; ZF_PHD_2; 1.
CC Nuclear protein; Zinc-finger.
CC ZN_FING 5 56 PHD-TYPE.
CC FT DOMAIN 236 336 JMC.
CC FT DOMAIN 472 518 PRO-RICH.
CC FT DOMAIN 487 648 LYS-RICH.
CC FT DOMAIN 960 1025 SER/THR-RICH.
CC FT CONFLICT 19 41 FMLECDACKDWHGSCGVGESEE -> PRAARPAPGPTR
AAQREGAT (IN REF. 2).

```
FT CONFLICT 75 76 QA -> PT (IN REF. 1).
FT CONFLICT 100 100 S -> R (IN REF. 2).
FT CONFLICT 106 106 A -> S (IN REF. 1).
FT CONFLICT 115 115 L -> V (IN REF. 1).
FT CONFLICT 621 621 K -> R (IN REF. 1).
FT CONFLICT 633 633 MISSING (IN REF. 1).
FT CONFLICT 642 642 K -> R (IN REF. 1).
FT CONFLICT 654 654 T -> S (IN REF. 1).
FT CONFLICT 664 664 N -> S (IN REF. 1).
FT CONFLICT 670 670 P -> A (IN REF. 1).
FT CONFLICT 698 698 N -> S (IN REF. 1).
FT CONFLICT 712 719 AYLPTPT -> BALMETS (IN REF. 1).
FT CONFLICT 727 727 A -> V (IN REF. 1).
FT CONFLICT 752 752 R -> K (IN REF. 1).
FT CONFLICT 782 782 S -> N (IN REF. 1).
FT CONFLICT 817 817 A -> T (IN REF. 1).
FT CONFLICT 823 823 S -> G (IN REF. 1).
FT CONFLICT 833 833 N -> I (IN REF. 1).
FT CONFLICT 837 841 SKKSA -> NMCT (IN REF. 1).
FT CONFLICT 849 849 A -> T (IN REF. 1).
FT CONFLICT 857 857 D -> E (IN REF. 1).
FT CONFLICT 862 862 MISSING (IN REF. 1).
FT CONFLICT 888 888 I -> V (IN REF. 1).
FT CONFLICT 949 950 SK -> NR (IN REF. 1).
FT CONFLICT 954 955 SA -> NT (IN REF. 1).
FT CONFLICT 959 960 LT -> PA (IN REF. 1).
FT CONFLICT 964 964 T -> A (IN REF. 1).
FT CONFLICT 968 968 T -> I (IN REF. 1).
FT CONFLICT 972 974 ISA -> ASAST (IN REF. 1).
FT CONFLICT 979 979 T -> A (IN REF. 1).
FT CONFLICT 1038 1040 TGA -> SGS (IN REF. 1).
FT CONFLICT 1045 1045 T -> A (IN REF. 1).
FT CONFLICT 1068 1068 MISSING (IN REF. 1).
SQ SEQUENCE 1101 AA; 121231 MW; A703F6C98C16B471 CRC64;
```

```
Query Match 5.6%; Score 167; DB 1; Length 1101;
Best Local Similarity 38.3%; Pred. No. 0.063;
Matches 31; Conservative 16; Mismatches 20; Indels 14; Gaps 4;

QY 269 LYCICQPHN-NRFMICCDRCBWFHGCVCISEARGRLLRNGEDYICPNCITILQVQDE 327
Db 6 VYCVCLPYDVTFRFMEICDACKDWFHGCVCVGEERBAPDI----DIYHCPNC-----EK 55

QY 328 THSETADQOEAKWR---PGDA 345
Db 56 THGKSTLKKRTWHRKGPQA 76
```

Search completed: April 30, 2004, 08:54:27
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 17:22:09 ; Search time 5297 Seconds
(without alignments)
4598.600 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGDPSNEEAPKAINKPTS.....RSFWIAIPWACPGIGVRAALC 562

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool_p/US09787016/runat 30042004 083702 4968/app query.fasta_1.711
-DB=GenEmbl -PMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016@cgn 1.1 5265 @runat 30042004 083702 4968 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2989	100.0	2610	6	BD244712 Genes enc
2	2989	100.0	2610	6	AX023369 Sequence
3	2989	100.0	2725	9	BC014489 Homo sapi
4	2989	100.0	2767	9	BC004237 Homo sapi
5	2981	99.7	2772	6	AX878224 Sequence
6	2981	99.7	2772	6	BD157094 Primer fo
7	2981	99.7	2772	9	AK002127 Homo sapi
8	2811	94.0	2407	9	BC000770 Homo sapi
9	2112.5	70.7	2602	10	MW238332 Mus muscu
10	2112.5	70.7	2867	6	BD244713 Genes enc
11	2112.5	70.7	2867	6	AX023370 Sequence
12	2081.5	69.6	4906	10	AK129117 Mus muscu
13	1575	52.7	6692	9	AB002331 Human mRN
14	1372.5	45.9	272404	2	AC108337 Rattus no
15	1342	44.9	104663	10	AL732560 Mouse DNA
16	1281	43.2	2428	5	BC060442 Xenopus 1
17	1282	42.9	160241	9	HS88561 Human DNA
18	1101	36.8	28519	9	HSU563E14 Human DNA
19	918.5	30.7	221547	2	AC119716 Rattus no
20	817.5	27.4	69252	2	AC101519 Mus muscu
21	549	18.4	681	6	AX869935 Sequence
22	549	18.4	681	6	BD149997 Primer fo
23	518	17.3	69252	2	AC101519 Mus muscu
24	465.5	15.6	198133	2	EX296530 Danio rer
25	419	14.0	2161	3	AK114522 Clona int
26	267.5	8.9	61204	2	AC017132 Drosophil
27	267.5	8.9	161601	3	AC007594 Drosophil
28	267.5	8.9	225655	3	AB003695 Drosophil
29	252	8.4	112387	2	AC098841 Magnapor
30	237.5	7.9	2567	9	HS880733 Homo sapi
31	232	7.8	5975	6	AX329603 Sequence
32	232	7.8	6256	6	AR338778 Sequence
33	232	7.8	6936	9	D87685 Human mRNA
34	232	7.8	6948	6	AX210673 Sequence
35	232	7.8	6948	9	AF091622 Homo sapi
36	232	7.8	7142	9	HS8808416 Homo sapi
37	231.5	7.7	2452	9	AF149758 Homo sapi
38	231.5	7.7	2487	9	AB031069 Homo sapi
39	228.5	7.6	2718	5	BC056775 Danio rer
40	214.5	7.2	2308	9	BC015733 Homo sapi
41	214.5	7.2	2320	6	AX086854 Sequence
42	214.5	7.2	2320	9	HS8801830 Homo sapi
43	214.5	7.2	2331	9	BC014940 Homo sapi
44	214.5	7.2	68050	8	EX842636 Neurospor
45	211	7.1	3298	8	AK111180 Oryza sat

ALIGNMENTS

RESULT 1

BD244712 2610 bp DNA linear PAT 17-JUL-2003
LOCUS Genes encoding for the human and murine death
DEFINITION inducer-obliterators-1.
BD244712
ACCESSION BD244712.1 GI:33054482
VERSION JP 2002526040-A/1.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2610)
AUTHORS Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, E. and Martinez, P.
TITLE Genes encoding for the human and murine death inducer-obliterators-1
JOURNAL Patent: JP 2002526040-A 1 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Homo sapiens (human)
PN JP 2002526040-A/1
PF 20-AUG-2002
PR 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5, 17-SEP-1998 US 60/100873 PI
CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC A61P35/00, A61K38/00, A61K45/00, A61K48/00, A61P3/10, A61P17/00, PC
A61P29/00,
PC A61P35/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12Q1/02, C12Q1/68, G01N33/15, G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00, A61K37/02
CC This gene is referred to in the application as Human Death CC
Inducer
CC Obliterators Gene 1.
CC It has now been named by the Human Gene
Nomenclature Committee
CC as
CC Death-Associated Transcription Factor (DATF-1) PH Key
Location/Qualifiers
FT source 1..2610
FT Location/Qualifiers
ORIGIN
1..2610
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
Alignment Scores:
Pred. No.: 1.22e-118 Length: 2610
Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-787-016A-3 (1-562) x BD244712 (1-2610)
QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
DB 265 ATGGACGACAAAGGCGACCGAGCAATGAGGAGCACTAAGGCCATCAACCCACAGC 324
QY 21 LysGluPheArgLysThrTroGlyPheArgArgThrLleAlaLysArgGluGlyAla 40
DB 325 AARGAGTTTCAGGAAACATGGGGTTTTCAGAGGACCATATTCGCCACGAGAGGGCCCA 384
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 385 GGGGACCGGAGGCTGACCCACTGGAGCGCCGCCACCCACAGCAGCAGCTGGGCTGTCC 444
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrile 80
DB 445 CTGGCGCGAGTGGGAGGCGCCCAAGCGCACTGAGCGCGTGGAGCACTTCTCTGACCA 504

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 505 GCGCGGCGCGCGCGAGGAGCATGCTCTCCCTGGAGGATCTGGTGGAGCCACG 564
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 565 TCTGCCCCGCCACAGAGCGCGAGACAGCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAG 624
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 625 ACCAGAGCGCGCGCGAGTCTGCTTCCACAGCTGTGTAAGGAACGACGAGCTCTTCTGAA 684
QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 685 AAGTGAAGAGGAGGAGTGCACGATGACACTCCGATAGTACAGCGATGGCTGAC 744
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysGluGlnGlnGluProGluPro 180
DB 745 TTGAAAGAGCTTCAGAAATCGCTTCGAGGAGCGGGAACAGGAGGCCACTGAGAGGCC 804
QY 181 LeuLysGlyLeuGlnSerArgLeuArgLysValArgArgGluGluGlyProAlaGluThr 200
DB 805 CTGAAGGGATCAGAGTCCGCTCCGGAAGAGCGCGGAGGAGGCTCCGCGGAGACT 864
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
DB 865 GTGGGCTCCGAGCGCAGTGCACCTGTGGAGGGGCTCTGCCCTAGTACGAGGAGCGCGAG 924
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
DB 925 AACGATCAGGGGGTGTGCTCCAGGCTGGGAAGATGACAGAGAGTAAATGGAGGGA 984
QY 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
DB 985 AAGCGGCTCAGGACATCAAGATGAGGCGCTGGAGACTTGGCGCGACCGAGCTGAA 1044
QY 261 CysGluGlyThrAspProAsnAlaLeuThrCysLysCysArgGlnProHisAsnArg 280
DB 1045 TGTGAGGGTTACGACCCCAACGCTTGTATTGCTATTCGCCAGCTCACAACAGG 1104
QY 281 PheMetileCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyLysSer 300
DB 1105 TTTATGATTTGCTGTGACCGCTGTGAGATGGTTTCATGCGATTCGTGGGCAATTCT 1164
QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrileCysProAsnCysThr 320
DB 1165 GAGGCTCGAGGAGGCTTTTGGAAAGGATGGGGAAGACTATATCTGCCCAAACTGCAC 1224
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
DB 1225 ATCTGCAAGTGCAGGATGAGATCAATTCAGAAACCGCAGATCAGCAGGAGCTAATGG 1284
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerileGlyThrileGluGlnLysSer 360
DB 1285 AGACCTGAGATGCTGATGCGACCGCTTGTACAGTATAGGACATATATCTGCCCAAACTGCAC 1344
QY 361 SerGluAspGlnGlyLysGlyArgLysGluLysAlaAlaAsnProSerGlyLysLys 380
DB 1345 AGCGAGACCAAGGAGTAAAGGGTAGAATTTGAGAAAGCTGCAATCCAGTGGCAGAG 1404
QY 381 LysLeuLysLysPheGlnProValLleGluAlaProGlyAlaSerLysCysLysGlyPro 400
DB 1405 AAACCTCAAGATCTTCCAGCTGTGATAGAGCGCGCTGGTGGCTCAAAATGTTATGGCCCC 1464
QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysLysLysLys 420
DB 1465 GGGTGTCTCAGTGGCGGAGCGGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1524
QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
DB 1525 CAGCGCGAGCGACATGAGATTTCTAAGCTCAGTAAAGACAGAGGCAAGCCCTAAA 1584

QY 441 GluLysMetIysMetIysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
DB 1585 GAAAGATGAAGATGAAGCCAGAGAGCCAGCTCTCCGAAATGCGGTCTCAGGCAGGT 1644
QY 461 IleIysIleSerSerValHisIysArgProAlaProGluLysGluThrThrValLys 480
DB 1645 ATTAAATCTCTCTGTGCAAGAGACAGCTCCAGAAAAAAGAGACCAAGTGAAG 1704
QY 481 LysAlaValValProAlaArgSerGluAlaLeuGluLysGluAlaAlaCysGluSer 500
DB 1705 AAGGCAGTGTCTCTCGCGAGTGAAGCACTCGGGAAGCAAGCAGCTTGTGAGAGC 1764
QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
DB 1765 AGCACCCGCTCGTGGCGAGCGATCAATTAATGCAATGCAATGCAATGCAATGCAAT 1824
QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
DB 1825 GCTCCCTCGCGCTCACTGTGTATTAATGATATATATATATATATATATATATATAT 1884
QY 541 ProSerArgSerPheThrIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
DB 1885 CCCTCCCGTCTCTTCTGGATAGCATCCCTCGGCGCTGTCCAGGACTGGGAGTTCAGCT 1944
QY 561 LeuCys 562
DB 1945 TTGTGT 1950
RESULT 2
AX023369 LOCUS 2610 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 1 from Patent WO0015787.
ACCESSION AX023369
VERSION AX023369.1 GI:10183781
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Leonardo, B.; Martinez, P.; Alonso, C.M.; Domingo, D.G. and Grandien, A.
TITLE Genes encoding for the human and murine death in
ducer-obliterators-1
JOURNAL Patent: WO 0015787-A 1 23-MAR-2000; CONSEJO SUPERIOR
LEONARDO ESTHER (ES); MARTINEZ PEDRO (ES); CONSEJO SUPERIOR
INVESTIGACION (ES); DOMINGO DAVID GARCIA (ES); MARTINEZ ALONSO
CARLOS (ES); BANNERMAN DAVID GARDNER (GB); GRANDIEN ALF (SE)
FEATURES
source
1. 2610
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="This gene is referred to in the application as
Human Death Inducer Obliterator Gene 1. It has now been
named by the Human Gene Nomenclature Committee as
Death-Associated Transcription Factor (DATF-1)."
ORIGIN
Alignment Scores:
Pred. No.: 1,22e-118 Length: 2610
Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-787-016A-3 (1-562) x AX023369 (1-2610)
QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
DB 265 ATGGACGACAAAGCGGACCGGAGCATGAGAGGACCTTAAGGCGCTCAAAACCCAGC 324
QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40

DB 325 AAGAGATTGAGGAAACATGGGGTTTTGAGAGGACCACTATCGCCAGGAGGGCGCA 384
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 385 GGGGAGCGGGAGGCTGACCCACTGAGCGGCCACCACACAGCAGCTGGGCTGTCTCC 444
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 445 CTGCGCGCGAGTGGAGGCGCCCAAGCGCATGTAGCGCGTGGAGCGGTTCCTGACCAT 504
QY 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 505 GGGCGCGCGCGCGGAGGAGGAGCATGCTGTCTCTCCCTGGAGGATTCGTGTGAGCCAGC 564
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 565 TCTGTCCCGCCACAGACGCCGAGACAGCTCCGAGGCGAGCGTGGAAAGCGCTTCTGAG 624
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 625 ACAGAGAGCGCGCCCGAGTCTGCTTCCACAGCTGTGNAGGACAGCAGCCCTTCTGAA 684
QY 141 LysValLysGlyGlyAlaAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 685 AAGGTGAAAGGAGGGGATGACCAAGATGACCTCCGATAGTCACAGCGATGCGCTGACC 744
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
DB 745 TTGAAGAGCTTCAGATCGCTTCCAGAGAGCGGAAACAGAGCCCTTCTGAGAGCGCC 804
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200
DB 805 CTGAAGAGGATCCAGAGTCTGCTCGGAGAGAGCGCGGAGAGGGTCCCGCCGAGACT 864
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerSerLysGlnProGlu 220
DB 865 GTGGGCTCCGAGCGCCAGTGACACTGTGGAGGGCGTCTGCCAGTAAAGCAGGAGCCGAG 924
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
DB 925 AAGCATCAGGGGTTGTGTCCAGGCTGGGAAGATGACAGAGAGAGTAAGTTGGAGGGA 984
QY 241 LysAlaAlaGluAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
DB 985 AAGGCGCTCAGGACATCAAGATGAGGAGCGCTGGAGACTTGGGCGGACCGAGCCTGAA 1044
QY 261 CysGluGlyTrpAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
DB 1045 TGTGAGGGTTACGACCCCAACGCCCTGTATTGCAATTTGCCGCGAGCTCACACACAGG 1104
QY 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
DB 1105 TTTATGATTGTCTGTACCGCTGTGAGAAATGGTTTCATGGCGATTGTGTGGGCAATTTCT 1164
QY 301 GluAlaArgGlyArgLeuLeuGluArgGluAspTyrIleCysProAsnCysThr 320
DB 1165 GAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGAGAGACTATATCTGCCCAACTGCACC 1224
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLysTrp 340
DB 1225 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGAGCTAAATGG 1284
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
DB 1285 AGACCTGGAGATGCTGATGGCCCGCATTTGTACAAGTATAGGAACAATAGACGAGAGTCT 1344
QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
DB 1345 AGCGAGACCAAGGATTAAGGGTAGAATTGAGAAAGCTGCAATTCGAAGTGGCAAGAAG 1404
QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400

Db 1405 AAACCTCAGATCTTCCAGCCTGTGATAGAGCGCCTGTGCTCAAAATGATTGGCCCC 1464

Qy 401 GlyCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysLeuLeuLys 420

Db 1465 GGGTGTGTCTCAGTGGCGCAGCCGACTCGGTGCTACTGCGAGTATGACTGTATCTCTCAA 1524

Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440

Db 1525 CACGCCGACGACCAATGAAGTTCTTAAGCTCAGGTAAAGACACAGACCAAGACTTAA 1584

Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460

Db 1585 GAAAGATGAAGATGAAGCAGAGAGAGCCAGTCTTCGGAATGCGGTCTCAGGAGGT 1644

Qy 461 IleLysIleSerSerValHisLysAspProAlaProGluLysLysGluThrThrValLys 480

Db 1645 ATTAATATCTTCTGTGCACAGAGACGAGCTCCAGAAAAAAGAGACCAAGTGAAG 1704

Qy 481 LysAlaValValProAlaArgSerGluAlaGluGlyLysGluAlaAlaCysGluSer 500

Db 1705 AAGCAGTGTGTGCTCTGCGCGAGTGAAGCACTCGGAGCAAGCACTCTGTGAGAGC 1764

Qy 501 SerThrProSerThrAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAla 520

Db 1765 AGCAGCGCGTGTGCGGAGCGAGTCAATACATGAGTAAAGCAGAAAGACTGCT 1824

Qy 521 AlaProSerProSerLeuLeuTyLysCysMetTyHisIleuGlyValGlyLeuLeuAsp 540

Db 1825 GCTCCCTCGCGTACTGTTGATAATGATATATATATATATATATATATATATATATAT 1884

Qy 541 ProSerArgSerPheThrIleAlaIleProThrAlaCysProGlyLeuGlyValAlaAla 560

Db 1885 CCTCCCGTCTTCTTGTGATGACCATCCCTCGGCTGTCCAGGACTGGAGTGTGACGT 1944

Qy 561 LeuCys 562

Db 1945 TTGTGT 1950

RESULT 3

BC014489

LOCUS

DEFINITION Homo sapiens death associated transcription factor 1, transcript variant 2, mRNA (cdna clone MGC:23216 IMAGE:4901057), complete cds.

ACCESSION BC014489

VERSION BC014489.1 GI:15680266

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2725)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Frange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Buffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, J.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

12477932

2 (bases 1 to 2725)

Strausberg, R.

Direct Submission

Submitted (17-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada

info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhur, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 34 Row: n Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

Location/Qualifiers

1. 2725

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:23216 IMAGE:4901057"

/tissue_type="Muscle, rhabdomyosarcoma"

/clone_lib="NIH MGC 17"

/lab_host="DH10B-R"

/notes="Vector: pOTB7"

1. 2725

/genes="DTF1"

/notes="synonyms: D101, FLJ11265, KIAA0333"

/db_xref="LocusID:11083"

/db_xref="MIM:604140"

238..1926

/codon_start=1

/product="death associated transcription factor 1, isoform a"

/protein_id="AAH14489.1"

/db_xref="GI:15680267"

/db_xref="LocusID:11083"

/translations="MDDKDPNNEAPKAKPTSKFRKTGFRRTTIARREGAGDAE ADPLEPPPOQGLSLRSGQPKHTEVEQFLATARRGSRMPVSLDSGEPTSC PADATASEGSVESASSTRSGQSNASTAVKRPASSEKVGKGDHDDTSDSDGLT LKELQNLKRKQREPTERPLKQISRLAKKRREGPAPTVSGEADTVGVLPSKQK FENDQGVSOAGKDDRESLEKGAQDIDKBPGLDGRPKPCBCEGDPNLYXCICQP HNNRFICDCRCBFWFHGDCVGISSEARGLLENGEDYICPNCTILQVQDTHSETAD QKAKWRPGDAGTCTCTGTTGKSSDQGIKRIKAANPSGKKKXIFQPVTEAP GASKICPGCHVAPQDSVVCNSDCLIKHAAATMKFLSSGKEQPKPKMKMKPEKP SLPKCAOAGIKISSVHKRPAPKKTIVKAVVVPARSEALGKKAACSSPTSPWASD HYNVAKPKTAAPSPLLYKVMHGVGLLDPSRSFWIAIPWACGLGVALLC"

1048..1197

/notes="PHD; Region: PHD-finger. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains"

/db_xref="CDD:pfam00628"

misc_feature

1.28e-118 Length: 2725

Alignment Scores:

Pred. No.:

Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-787-016a-3 (1-562) x BC014489 (1-2725)

QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaLeuLysProThrSer 20
DB 238 ATGGACACAAAGCGGACCGCCGAGCAATGAGGAGCCACTTAAGGCCATCAAAACCCACGAGC 297
QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
DB 298 AAGAGTTTCAAGAAACATGGGTTTTCGAAGGACCACTATCGCCAGCGAGAGGGCGCA 357
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
DB 358 GGGGACCGGAGGCTGACCCACTGGAGCGCCACCCGACGAGCAGCGAGCTGGGCCCTGTCC 417
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 418 CTGGCGCGCAGTGGAGGCGCCAGCCAGCCACTGAGCGCGCTGAGCGAGTTCTTCGACCAT 477
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 478 GCGCGCGCGCGGAGGAGCATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 538 TCTTGCCTCCGACACGACGCGACAGCTCTCGAGGCGAGCTGAGGAGCGCTTCTGAG 597
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 598 ACCAGAGCGCGCCCGCCAGTCTGCTTCCACAGCTGTGAGGAGAACGACGAGCTCTTCTGAA 657
QY 141 LysValLysGlyLysAspAsnHisAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 658 AAGGTGAAGAGGAGGAGTACCCAGTATGACACCTCCGATAGTACAGCGATGGGCTTGACC 717
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
DB 718 TTGAAAGAGCTTCAGATCGCTTTCGAGGAGCGGACGAGGAGCGCCACTGAGAGGCGCC 777
QY 181 LeuLysGlyLysGlnSerArgLeuArgLysArgGluGlnGluGluProAlaGluThr 200
DB 778 CTGAAGAGGATCCAGAGTCCGCTCGGAGAGAGCGCGCGGAGGAGGCTCCCGCGAGACT 837
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
DB 838 GTGGGCTCCGAGGCGGAGTACACTGTGAGGAGCGCTCTGCGGAGTACAGGAGGAGCGAG 897
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
DB 898 AACGATCAGGGGGTGTGTGCTCCAGGCTGGGAAAGATGACAGAGAGATGATGAGTGGAGGGA 957
QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
DB 958 AAGCGCGCTCAGGACATCAAGATGAGGAGCTTGAGACTTGGGCGCCAGGAGGAGCTGAA 1017
QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
DB 1018 TGTGAGGTTTACGACCCCGCCCTGTATTGCTATTTGCGCCGAGCTTCAACACAGG 1077
QY 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSer 300
DB 1078 TTTATGATTTGCTGTGACCGCTGGAAGAATGGTTTATGCGGATGTGTTGGGCGATTTCT 1137
QY 301 GluAlaArgGlyArgLeuGluGluArgAsnGlyGluAspFyrIleCysProAsnCysThr 320
DB 1138 GAGGCTTCAGGGAGGCTTTTGGAAAGGATGGGAGAGCTATATCTGCCCAACATGCAAC 1197
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340

DB 1198 ATTCTCAGTGCAGGATGAGACTCATTCCAGAAACCGCAGATCAGCAGGAGCTAATGG 1257
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
DB 1258 ACACCTGGAGATGCTGATGCGCCGATTTGTAAGTATAGGACCAATAGCAGGAGTCT 1317
QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
DB 1318 AGCGAGACCAAGGATTAAGGTAGAAATGAGAAAGCTGCAATCCAAAGTGGCAAGAAG 1377
QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
DB 1378 AAACTCAGATCTTCCAGCTGTGATAGAGCGCCCTGGTGCCTCAAAATGTATTGGCCCC 1437
QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
DB 1438 GGTGCTGTCTGTCGCGCAGCCGACTCGGTACTCTGAGTAATGACTGTATCTCTCAA 1497
QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 440
DB 1498 CACGCGCAGCGACATGAAGTTTCTTAGCTCAGGTAAAGAACAGAGGCAAGAGCTTAA 1557
QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
DB 1558 GAAAGATGAGATGAAGCAGAGAGCCGAGTCTTCCGAAATGCGGTCTCAGGAGGT 1617
QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
DB 1618 ATTAATAATCTCTCTGTCGACAGACAGCAGCTCCAGAAAAAAGAGAGCAGTGAAG 1677
QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
DB 1678 AAGCGAGTGGTGTCTCTCGCGAGTGAAGCAGCTCGGAGAGGAGCAGCTTGTGAGGC 1737
QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
DB 1738 AGCAAGCGCTCTGCGGAGGAGTCAATTAATGAGTAAAGCAGAGAGAGAGCTGCT 1797
QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
DB 1798 GCTCCCTCCGCGCTCCTGTGTATTAATGTATGATCCTAGGCGGTGGCCCTCCGAGC 1857
QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
DB 1858 CCTCCCGTCTCTCTGATAGCCATCCCTGGGCTCTCCAGGACTGGGAGTTGAGCT 1917
QY 561 LeuCys 562
DB 1918 TTGTGT 1923

RESULT 4

BC004237

LOCUS

DEFINITION

Homo sapiens death associated transcription factor 1, transcript

variant 1, mRNA (cDNA clone MGC:896 IMAGE:2959493), complete cds.

ACCSSION

BC004237.2

VERSION

MGC.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 2767)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klaunig, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zensberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, I., Max, S.I., Wang, J., Hong, P.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalls, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2767)
Straussberg, R.
Direct Submission
Submitted (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:13278980.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadansystemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whitting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 2 Row: 0 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18375620.

FEATURES

Source

1..2767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:896 IMAGE:2959493"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH MGC 15"
/lab_host="PH10B-R"
/note="Vector: pOTB7"
1..2767
/gene="DATF1"
/note="synonyms: DIO1, FLJ11265, KIAA0333"
/db_xref="LocusID:11083"
/db_xref="MIM:604140"
282..1970
/codon_start=1
/product="Death associated transcription factor 1, isoform
a"

gene

CDS

/protein_id="AAH04237.1"
/db_xref="GI:13278981"
/db_xref="LocusID:11083"
/translation="MDDKDPNSBAPKALPTSKFRKTKWFRFTTIKRGAGDAE
ADLPBPPOOGLSLRSRGPKEVERVDFPLIARRGERSNPSVLESDGPTSC
PATDAETASGVESASSTRSPQASNAVERPASSESRYKGGDDHDTSDSDSLT
LKEQLNRKKEQSPRTPLKIQSLRKKRRRGPAPATVGSSEADSTVEGVLPKQE
PENNQVWSQAGDRKLEKFAAQIDNBPEDGLRPKECEGYDPNLYICRQP
HNNRPMICDRCEWFGDCVIGISBARGLLERNEDYICFNCTILQVDETHSTAD
QGEAKWRPGDCCHVAPDSVYCSNDICILKHAATMKFLSSGKEQKPKPKMKMKPKP
GASKICIGKACCHVAPDSVYCSNDICILKHAATMKFLSSGKEQKPKPKMKMKPKP
SLPKGAQAGIKLSVKEPAPKPKETTVKAAVVVPAPSEALGKAEACESTPSWASD
HNNRPMKTRAPSPSLYKCMYHLGVGLLDPSPSFNIALPWACPGLVGAALC"
1092..1241
/note="PHD; Region: PHD-finger. PHD folds into an

misc_feature

interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"

ORIGIN

Alignment Scores:

Pred. No.: 1..3e-118 Length: 2767
Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x BC004237 (1-2767)

Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLysProThrSer 20
Db 282 ATGACACCAAGCGGACCCGAGCAATGAGGAGGACCTAAGGCCATCAACACCCACGAGC 341
Qy 21 LysGluPheArgLysThrTgPlyPheArgThrThrileAlaLysArgGluGlyVala 40
Db 342 AAAGAGTTTCAGGAAACATGGGGTTTTCGAAGGACCACTATGCGCAAGCGAGAGGCGCA 401
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 402 GGGGACGCGGAGGCTGACCCACTGAGCGCGCCACCCCCACAGCAGCAGCTGGGGCTTGTCC 461
Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrile 80
Db 462 CTGGGGCGCAGTGGAGGAGCGCCCAAGCGCACTGAGCGCGTGGAGCAGTCTCTGACCAT 521
Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 522 GCGCGGCGCGCGGAGGAGGAGCATGCTGTCTCTCTCTGGAGGATCTCTGTGAGCCACG 581
Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 582 TCCTGCCCCCGCCACAGACCGCCGACAGCCTCCGAGGCGCGCTGGAAAGGCGCTTCTGAG 641
Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 642 ACCAGAGCGGCGCCCGCAGTCTGCTTCCACAGCTGTGAAGGAACACGACGAGCCCTCTTCTGA 701
Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 702 AAGGTGAAGAGGAGGAGGATGACCAAGTACACCTCCGATAGTACAGCGATGAGCTGAGC 761
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGlnProThrGluArgPro 180
Db 762 TTGAAGAGCTTCAGATCGCTTCCAGGAAGCGGGAACAGGAGGCCCACTGAGAGGCGCC 821
Qy 181 LeuLysGlyLysGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
Db 822 CTGAAGGAGATCCAGAGTCGCTCGCGAAGAGGCGCGGAGGAGGCTCCCGCGAGACT 881
Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 882 GTGGGCTCCGAGGCGAGTGAACACTGTGGAGGCGCTCTCTCCCACTAGCAGAGGCGCGAG 941
Qy 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Db 942 AACGATCAGGGGCTGTGTGTCAGCTCGGAGGAAAGATGACAGAGAGAGTAGTAACTTGAGGGA 1001
Qy 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db 1002 AAGCGCGCTCAGGACATCAAGATGAGGAGGCTTGAGACTTGGCGCGGAGCGGAGCGCTGAA 1061
Qy 261 CysGluGlyTyAspProAsnAlaLeuTyrCysLysCysArgGlnProHisAsnArg 280
Db 1062 TGTGAGGGTTACGACCCCAAGCGCTGTATTGCACTTTCGCCCGCAGCTCTCACACACAGG 1121
Qy 281 PheMetLeuCysCysAspArgCysGluGluThrPheHisGlyAspValGlyLysSer 300

Db 1122 TTTATGATTGCTGTGACCGCTGTGAGGAATGGTTTTCATGGCAATGTGTGGCGATTTCT 1181
 Qy 301 GluAlaArgGlyValGluLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 Db 1182 GAGGCTCGAGGGAGGCTTTTGGAGAGATGGGAGAGATATATCTGCCCAACTGCAC 1241
 Qy 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaValTyr 340
 Db 1242 ATTCTCAAGTCAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGGAAGCTAATGG 1301
 Qy 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 Db 1302 AGACCTCGAGATGCTCATGACCGAGTGTACAGTATAGGACATATAGCAGCAGAGTCT 1361
 Qy 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 Db 1362 AGCGAAGACCAAGGATAAAGGTAGATTTGAGAAAGCTGCRAATCAAAGTGGCAGAG 1421
 Qy 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1422 AAATCAAGATCTTCAGGCTGTGATAGAGCGCCCTGGTGCCTCAAAATGTATGGCC 1481
 Qy 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 Db 1482 GGTGTGTGTACGTGGCGAGCCCGACTCGGTGTACTGTGAGTATGACTGTATCTCAA 1541
 Qy 421 HisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1542 CACGCGCGACGACAAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAAGCCAAAGCTAAA 1601
 Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1602 GAAGAATGAGATGAGCCAGAGAGCCCGACTCTCCGAAATGCCGTCTCAGGAGGT 1661
 Qy 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1662 ATTAATATCTTCTGTGCACAAAGAGACCACTCCAGAAAAAAGAGACCAAGTGAAG 1721
 Qy 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 Db 1722 AAGCGAGTGTGTCTTCTGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGGC 1781
 Qy 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 Db 1782 AGCACGCGTGTGGCGAGCGATCAATTAATCAATGAGTAAAGCCAGAAAGACTGCT 1841
 Qy 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 Db 1842 GCTCCCTCGCGCTCACTGTGTATAATGTATGTATCATCCTAGGGGTGGCCCTCTCGGAC 1901
 Qy 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 Db 1902 CCTCCCGTCTTCTTGATAGCATCCCTCGGCGCTGTCCAGACTGGGAGTTCAGCT 1961
 Qy 561 LeuLys 562
 Db 1962 TTGTGT 1967
 RESULT 5
 AX878224
 LOCUS AX878224 2772 bp DNA linear PAT 17-DEC-2003
 DEFINITION Sequence 13129 from Patent EP1074617.
 ACCESSION AX878224
 VERSION AX878224.1 GI:40032960
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 TITLE Primers for synthesising full-length cDNA and their use

JOURNAL Patent: EP 1074617-A 13129 07-FEB-2001;
 Research Association for Biotechnology (JP)
 FEATURES
 source
 1. 2772
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 301.1989
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAE90061.1"
 /db_xref="GI:40032961"
 /translation="MDDKGDPSNEBAPKAIKPTSKBRTKWTGFRRTTIAKREGADAE
 ADLPPPPQQLGLSLRSGRQKRTVERVQFLTIARRGRSRVSLDSGEPTSC
 PATTAETASGEVSASSTSGPOASTAVKERPASSEKVKGGDDHDDTSDSDGLT
 LKELQNRKRREORPTERPLKIQSLRKKRRERGPATVGESEADTVGVLPKQK
 PENDINGVVSAGKDDREKLGKAQDIKDEPREDLGRPKCEKGVDPNALYICROP
 HNNRPMICDRCEWFHDCVGLSEARGILLERNGEDVCPNCTILOQDEPESFAD
 QQEAQRGDDADGTDCTSIGTILEEQSSDDQIKGRIKAAANPSGKKKLKIFQVTHAP
 GASKICGCGCHVAPDSVYSCNDCLIKHAAATMFLSSGRKQPKPKMKMKQKPKP
 SLPKCGAGGAKISSVHRPAPKCTTTKCAVVVPAVSEALGKAAACSSPTPSWASD
 HNTYAVKPEKTAAPSPSLLYKMYLHGLVGLDPSRSPWIAIPWACPGGLGVAALC"
 ORIGIN
 Alignment Scores: 2.84e-118 Length: 2772
 Pred. No.: 2981.00 Matches: 561
 Score: 99.82% Conservative: 0
 Percent Similarity: 99.82% Mismatches: 1
 Best Local Similarity: 99.73% Indels: 0
 Query Match: 6 Gaps: 0
 DB: 0
 US-09-787-016A-3 (1-562) x AX878224 (1-2772)
 Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 301 ATGACGACGACAAAGGGGACCCGAGCAATGAGGAGCCACTAAGCCCATCAACCCACGAG 360
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
 Db 361 AAAGAGTTTCAGGAATCAATGGGGTTCGAGAGCCACTATCGCCCAAGCGAGGCGCGCA 420
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 421 GGGGAGCGGAGGCTGACCCACTGAGGAGCCGCCACCCACAGCAGCAGCTGGGCTGTCC 480
 Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 481 CTGCGCGCGCAGTGGGAGGCGAGCCCAAGCGCACTGAGCGCTGGAGCAGTTCCTGACATT 540
 Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 541 GCGCGCGCGCGGCGAGGAGGAGCATGCTGTCTCTCTGGAGGATTCCTGAGGCGCCAG 600
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 601 TCTCTGCGCCGACAGAGCCGAGAGCAGCTCCGAGGGCAGCGTGGAAAGCGCTTCGAG 660
 Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 661 ACCAAGCGCGCGCGCGGAGGAGGAGCATGCTGTCTCTGAGGAGGAGGAGGAGGAGG 720
 Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 721 AAGGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
 Qy 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGlnGluProThrThrGluArgPro 180
 Db 781 TTGAAGAGGAGTTCAGAAATGCGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
 Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluGluGluProAlaGluThr 200
 Db 841 CTGAAGGAGGATCCAGATGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900

QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 901 GTGGGCTCCGAGGCGAGTGCACCTGTGAGGCGCTCTCTGCCACCTAGCAGGAGCCGCGAG 960

QY 221 AsnAspGlnGlyValValSerGlnAlaGlyValSerAspArgGluSerLysLeuGlnGly 240
 Db 961 AACCATCAGGGGGTGTGTCCAGCGCTGGGAAAGATGACGAGAGAGTAAGTTGGAGGGA 1020

QY 241 LysAlaAlaGlnAspIleLysAspGlnGluProGlyAspLeuGlyArgProLysProGlu 260
 Db 1021 AAGCGGCTCAGGACATCAAGATGAGGAGCGCTCAAGACTTGGCGCCGACCGAGCCTGAA 1080

QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 Db 1081 TGTGAGGGTTACGACCCCAAGCCCTGTATTGCAATTTGCCGCCAGCCTCACAAACAGG 1140

QY 281 PheXetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300
 Db 1141 TTTATGATTTCTGTGACCGCTGTGAGAAATGGTTTCATGGCGATTTGTGGCGATTCT 1200

QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 Db 1201 GAGGCTCAGGAGGCTTTTGGAAAGGAATGGGAAGACTATATCTGCCCAACTGCACC 1260

QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyr 340
 Db 1261 ATTTCTGCAAGTGCAGATGAGACTCATTCAGAAACGGCGATTCAGCAGGAGCTAATGG 1320

QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 Db 1321 AGACCTGAGATGCTGATGGCACCGATTGTACAGTATAGGAACAATAGAGCAGAAAGTCT 1380

QY 361 SerGluAspGlnGlyIleGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 Db 1381 AGCGAAGACCAAGGATAAAGGTAGAAATTGAGAAGCTGCATTCCTCANGTGGCAGAG 1440

QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1441 AAACTCAAGATCTTCCAGCTGTGTATAGAGGCGCTGTGTCTCAAAATGTATTGGCCCC 1500

QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 Db 1501 GGGTCTGTCTGCTGCGCGAGCCGACCTCGGTGACTGCGATGATGATGATCTCTCAAA 1560

QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1561 CACGCCGCGAGCAGCANTGAGTTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCCTAAA 1620

QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1621 GAAAAGATGAATGAAGCAGAGAGCCAGCTCTTCCGAAATGCGGTCTCAGGACGT 1680

QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1681 ATTAAATCTTCTGTGTCACAGAGACCCAGCTCCAGAAATAAAGAGACACAGTGAAG 1740

QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 Db 1741 AAGGCGAGTGTGCTCCCTGCGCGAGTGAAGCACTCGGAGAGAGAGAGCTGTGTGAGAGC 1800

QY 501 SerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 Db 1801 AGCAGCGCTGTGCGCGAGGAGGATCACATTTAAATGATGATGATGATGATGATGATGAT 1860

QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 Db 1861 GCTCCCTGCGCGCTACTGTTGTATATAATGATATGATATCACCCTAGGCGTTGGCCTCCTGAC 1920

QY 541 ProSerArgSerPheThrIleAlaIleProTyrAlaCysProGlyLeuGlyValAlaAla 560
 Db 1921 CCTCCCGCTTCTTCTGTGATAGCCATCCCTCTGGGCTGTCTCAGGAGTGTGAGTTCAGCT 1980

QY 561 LeuCys 562
 Db 1981 TTGTGT 1986

RESULT 6
 BD157094
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BD157094 2772 bp DNA linear PAT 17-JAN-2003
 Primer for synthesizing full-length cDNA and use thereof.
 BD157094.1 GI:27862852
 JP 2002191363-A/11937.
 Homo sapiens (human)
 Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 Primer for synthesizing full-length cDNA and use thereof
 HELIX RESEARCH INSTITUTE
 PATENT: JP 2002191363-A 11937 09-JUL-2002;
 OS Homo sapiens (human)
 PN JP 2002191363-A/11937
 PD 09-JUL-2000
 PF 28-JUL-2000 JP 2000280990
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
 PI SAITO,
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
 PI KEIICHI NAGAI, TETSUJI OTSUKI
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10,
 PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key
 Location/Qualifiers
 (301) (1986).

FT CDS Location/Qualifiers
 source
 i. 2772
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores: 2.84e-118 Length: 2772
 Pred. No.: 2981.00 Matches: 561
 Score: 99.82% Conservative: 0
 Percent Similarity: 99.82% Mismatches: 1
 Best Local Similarity: 99.73% Indels: 0
 Query Match: 6 Gaps: 0
 DB:

US-09-787-016A-3 (1-562) x BD157094 (1-2772)

QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
 Db 301 ATGACGACAAAGGCGACCCGAGCAATGAGGAGGACCTAAGGCCATCAACCCACCAGC 360

QY 21 LysGluPheArgLysThrTyrGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 Db 361 AAACAGTTTCAGAAAAAATGCGGGTTTTCGAGGACCACTATCGCCAGAGGAGGCGCA 420

QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 421 GGGGACGCGGAGGCTGACCCACTGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTGCC 480

QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 481 CTGGGAGCGCATGTGGAGGCGAGCCCGACGACCTGAGCGCGTGGAGGAGCTTCTGACCAT 540

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 541 GCGGCGCGCGCGGACGAGGAGGAGCATGCTGTCTCTCCCTGGAGGATTTCTGGTGGAGCCACG 600

101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db |TCTGCCCCCGCACAGACCGCGAGACAGCTCCGAGGGCGAGCGTGGAAAGCGCTTCAG 560
121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db |ACCAGAAAGCGGCGCCCGAGTCTGCTCCACAGCTGTGAAGGACGACACAGCTCTTCTGAA 720
141 LysValLysGlyCysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db |AAGGTGAAGAGGAGGAGATGACACAGATGACCTCCGATAGTGAACGATGCGCTGAC 780
161 LeuLysGluLeuGlnAsnArgLeuArgArgGluGluGlnGluProThrGluArgPro 180
Db |TTGAAGAGCTTCAGAAATCGCTTCGACGAAAGCGGGAACAGGAGCCCACTGAGAGGCC 840
181 LeuLysGlyLeuGlnSerArgLeuArgLysArgArgGluGluGluGluProAlaGluThr 200
Db |CTGAAGAGGATCCAGAGTCCGCTCGCGAAGAGCGCGGAGAGGCTCCCGCGAGACT 900
201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db |GTGGGCTCCGAGCCAGTGACACTGTGGAGGGGTCTGCCCACTAAGCAGGAGCCGAG 960
221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Db |AAGCATCAGGGGCTTGTGTCACAGGCTGGGAAAGATGACAGAGAGTAAAGTTGGAGGGA 1020
241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db |AAGCGGCTCGACATCAAGATGAGAGCTGGAAGACTTGGCGCCGACCGAGCTGAA 1080
261 CysGluGlyPyrAspProAsnAlaLeuPyrCysIleCysArgGlnProHisAsnAsnArg 280
Db |TGTGAGGTTTACGACCCCAACGCGCTGTATTGCAATTTGCCGCGCAGCTTCACAAACAGG 1140
281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyLysSer 300
Db |TTTATGATTGCTGTGACCCCTGTGAAGATGGTTTCATGGCGATTTGTGTGGCATTTCT 1200
301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyCysAspPyrIleCysProAsnCysThr 320
Db |GAGCTCGAGGGAGGCTTTTGGAAAGGATGGCGAAGACTATATCTGCCCAACTGCACC 1260
321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
Db |ATTCTGCACTGAGGATGAGCTCATTCAGAAACGCGCAGATCAGCAGGAAGCTTAATGG 1320
341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
Db |AGACCTGGAGTCTGTATGTCACCGATTGTACAAGTATAGAAACATAGAGCAGAAAGTCT 1380
361 SerGluAspGlnGlyLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
Db |AGCGAAGCAAGGAGATAAAGGGTAGAATTTGAGAAGCTCCAAATCCAAAGTGGCAAGAG 1440
381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
Db |AAACTCAAGATCTCCAGCTGTGATGAGAGCGCTGTGCTCCCAAAATGATTTGGCCCC 1500
401 GlyCysCysHisValAlaGlnProAspSerValPyrCysSerAsnAspCysIleLeuLys 420
Db |GGGTGCTGTCACTGGGCGAGCCGACTCGGTGACTGCAAGTAACTGATCTCTCTCAA 1560
421 HisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
Db |CAGCCCGCAGCGACAAATGAAGTTTCAAGCTCAGGTAAAGACAGAGCCAAAGCCCTTAA 1620
441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyValAlaGlnAlaGly 460
Db |GAAAGATGAGATGAGAGCCAGAGCCAGCTCTCCGAAATCGGCTGTCCAGCAGGT 1680
461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrValLys 480

1681 ATTAATCTCTCTGTGCACAGAGACCGCTCCAGAAAAAGAGACCAACGTGAAG 1740
481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
Db |RAGGCGATGGTGCTCCCTCGCGAGTGAACACTCGGAGAGGAGCAGCTTGTGAGAGC 1800
501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
Db |AGCAGCGCTGTGTCGCGAGCGATCACTATCAATGCAATGCAATGCAATGCAATGCAATG 1860
521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
Db |GTCCTCTGCGCTCACTGTGTATTAATATGATATCATCCTAGGAGTGGCTCTCTGGAC 1920
541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
Db |CCCTCCGCTCTCTTCTGATAGCCATCCCTGGGCTGTCCAGGACTGGGAGTTGAGCT 1980
561 LeuCys 562
1981 TTGTGT 1986

RESULT 7
AK002127 2772 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ11265 fis, clone PLACE1009158.
ACCESSION AK002127
VERSION AK002127.1 GI:7023814
KEYWORDS cDNA capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2772)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers
1. .2772
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1009158"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"
301. 1989
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA92094.1"
/db_xref="GI:7023815"
/translation="MDDKGDPSNEAPKAKFTSKFRKTGWRRTTIARREGAGDAE
ADPLEPPQQQLGLSLRSGRQPKTERVEQPLTIARRRRGRSMFVLSDSGPTSC

FEATURES
source
1. .2772
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1009158"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"
301. 1989
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA92094.1"
/db_xref="GI:7023815"
/translation="MDDKGDPSNEAPKAKFTSKFRKTGWRRTTIARREGAGDAE
ADPLEPPQQQLGLSLRSGRQPKTERVEQPLTIARRRRGRSMFVLSDSGPTSC

PATDAETASGVSASSETRSQPOSASTAVKPERPASSEKVKYGGDDHDDTSDSDSGLT
 LKLEONLRREOREQPTERPLKIQSRLRKKREBEPAETVGSASDTEVGLVPSKOE
 PENQGVVYQAGKDDREKSLKAQADIKLEBEDLGRPKPECEGYDPNLYCICROP
 HNNRPMICDCEEFHEDGCVGISSEARGLLERNGEDVICPNCETILOVDTHSETAD
 QOAKWRPDAADGDTCTSIGIHEKSSDQIKRIIEKANNPSCKKLKLPQVLEAP
 GASKCITGECCHVAPDSVYICNSDCILKRAAATKFLSSGREQKPKERKKMKDEPK
 SLPKGQAGAGIKISVHRPAPKETTIVKAAVVVPKSEALGKEAKESSTPWSAD
 HNYNAVPEKTAAPSPSLLYKQVHLGVLDPSPSEFWIALPMACPGLVNAALC"

ORIGIN

Alignment Scores:

Pred. No.: 2,848-118 Length: 2772
 Score: 2981.00 Matches: 561
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.73% Indels: 1
 Ds: 9 Gaps: 0

US-09-787-016a-3 (1-562) x AK002127 (1-2772)

Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 301 ATGGACGACAAAGGCGACCCGAGCAATGAGGAGGACCTTAAGGCCATCAACCCACCCAGC 360
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGluVala 40
 Db 361 AAAGAGTTTCAGGAACAACATGGGGTTTCGAGAGCACTATGCCACAGCAGAGGCGCA 420
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 421 GGGGACGGGAGGCTGACCACTGGAGCGCGCCACCCACACAGCAGCAGCTGGGGCTGTCC 480
 Qy 61 LeuArgArgSerLysArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 481 CTGGCGCGCAGTGGGAGGAGCGCCAAAGCGCATCTGAGCGGTGGAGCAGTTCTGACCAFT 540
 Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 541 GCGGCGCGCGCGGAGGAGGAGCATGCTGCTCTCCCTGAGGATTTCTGGTGAGCCACG 600
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 601 TCTGTCGCCCGCACAGACCCGAGCAGCTCCGAGGCGCGGTGGAAAGCCCTTCTGAG 660
 Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 661 ACCAGAGCGGCCCCCAGCTCTGCTTCCAGCTGTGAGAGACGACCCCTCTTCTGAA 720
 Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 721 AAGGTCAAGAGGAGGAGGATGACCATGACATCCGATAGTGACAGCGATGGCCTGACC 780
 Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 Db 781 TTGAAAGAGCTTCAGAAATCGCTTCGAGAGAGCGGGAAACAGGAGCCCACTGGAGGCGCC 840
 Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysArgArgGluGluGluGlyProAlaGluThr 200
 Db 841 CTGAAAGGATCCAGAGTCCCTGCGGAAGAGCGCGCGAGGAGGCTCCCGCCGAGACT 900
 Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 901 GTGGGCTCCGAGGCCCATGACATGTGGAGGGGCTCTGCCCATGTAGCAGAGAGCCCGAG 960
 Qy 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 Db 961 AACGATCAGGGGCTGTGTGTCCTCCAGGCTGGGAAAGATGACAGAGAGTAGTAAAGTTGAGGCA 1020
 Qy 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 Db 1021 AAGCGCGCTCAGACATCAAGATGAGGAGCCCTGAAGACTTGGCGCCGAGCGAGCTTGAA 1080
 Qy 261 CysGluGlyTyTrAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArg 280

Db 1081 TGTGAGGGTTACGACCCCAACGCCCTGTATGTGATTTTCCGCCAGCCTCACACACAGG 1140
 Qy 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
 Db 1141 TTTATGATTTGCTGTGACCGCTGTGAAGAAATGGTTTTCATGCGAFTGTGTGGCATTTCT 1200
 Qy 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThr 320
 Db 1201 GAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAACTGCACC 1260
 Qy 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLysTrp 340
 Db 1261 ATTCTGCAAGTGCAGATGAGACTCATTCAGAAACCGCAGATCAGCAGGAAGCTAAATGG 1320
 Qy 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlyLeuSer 360
 Db 1321 AGACCTGGAGATGCTGTATGGCAGCCGATTGTACAAATATAGAAACAATAGACCAAGTCT 1380
 Qy 361 SerGluAspGlnGlyIleLysGlyValGlyIleGluLysAlaAlaAsnProSerGlyLysLys 380
 Db 1381 AGCGAAGACCAAGGGATTAAGGTAGAAATTTGAGAAAGCTCCANATCCAAAGTGGCAAGAG 1440
 Qy 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1441 AAACCTCAAGATCTTCCAGCCTGTGATAGAGCGCCTGCTGCTCCTCAAAATGTATTGGCCCC 1500
 Qy 401 GlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLys 420
 Db 1501 GGGTGTCTGTCTACGTGGCGCGCCGCGCTCGGTGTATCTGCGAGTATGACTGTATCTCTCAA 1560
 Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1561 CACGCCGCGAGCACAAATGAAGTTTCTAAGCTCAGTAAAGAAACAGAAAGCCAAAGCCTAAA 1620
 Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyValaGlnAlaGly 460
 Db 1621 GAAAGATGAAAGTAAAGTAAAGCCAGAGCCAGTCTTCGAAATTCGCGTCTCAGCAGGT 1680
 Qy 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1681 ATTAATCTCTCTGTGTGACACAGAGACCCAGCTCCAGAAAAAAGAGACACACAGTGAAG 1740
 Qy 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 Db 1741 AAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1800
 Qy 501 SerThrProSerTipAlaSerAspHisAsnTyTrAsnAlaValLysProGluLysThrAla 520
 Db 1801 AGCAGCCCTCTGTGGCGCGGATCACAATTCAGATGCAATGCAAGTAAAGCCAGAAAGACTGCT 1860
 Qy 521 AlaProSerProSerLeuLeuTyIlyCysMetTyHisLeuGlyValGlyLeuLeuAsp 540
 Db 1861 GCTCCCTCGCGTCACTGTGTATAAATGTATGTATATCACTAGGGGTGGGCTCTCTGGAC 1920
 Qy 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 Db 1921 CCTCCCGCTTCTTCTGTGATAGCATCCCTGGGCTCTGCCAGTCCAGGAGTGGAGTTCAGCT 1980
 Qy 561 LeuCys 562
 Db 1981 TTGTGT 1986

RESULT 8

BC000770

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC000770 2407 bp mRNA linear PRI 12-NOV-2003
 Homo sapiens death associated transcription factor 1, transcript
 variant 3, mRNA (cdna clone MGC:3257 IMAGE:3506207), complete cds.
 BC000770
 BC000770.2 GI:33990873
 MGC
 Homo sapiens (human)
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2407)

Strausberg R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D., Altshuler, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.B., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McGowan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smalys, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22389257

12477932

2 (bases 1 to 2407)

Strausberg, R.

Direct Submission

Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Aug 20, 2003 this sequence version replaced gi:12653952.

Contact: MGC help desk

Email: cgabp@remail.nih.gov

Tissue Procurement: DCTD/BTP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: amad@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 7 Row: 1 Column: 2

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

Location/Qualifiers

1..2407

/organism="Homo sapiens"

/mol_type="cDNA"

/db_xref="taxon:9606"

/clone="MGC:3257 IMAGE:3506207"

/tissue_type="Kidney, renal cell adenocarcinoma"

/clone_lib="NIH MGC_14"

/lab_host="DH10B-R"

/note="vector: pOTB7"

1..2407

/genes="DAFP1"

/notes="synonyms: DI01, FLJ11265, KIAA0333"

/db_xref="LocusID:11083"

/db_xref="MIM:604140"

201..1835

/codon_start=1

/product="death associated transcription factor 1, isoform b"

/protein_id="AAH00770.1"

/db_xref="GI:12653953"

/db_xref="LocusID:11083"

/translation="MDDKGDPSNEEAPKAIPTSKFRFTWGRFTTIKREGAGDAEADPLEPPPPQOGLSLRSRQPKRTVERVOFLIARRGRSRMPSVLEDSGEPTSCPATDAETASGEVSASSTRSQSASAVKRPASSEKVGDDDDHDDTSDSDGLTAKELONLRARRKREBPTEPLKGIOSRLRKRRRERPARTVGSSEADTVGVLPKSOEPENDQGVVSQAGDDREKLEKKAQDIDKDEPDLGPKPECEGYDNALVCIQEQENRFRICDRCRWFHGDVCSISARGLRLERNGEDYICPNTCIYLVQDETHSTADQORAKRWDADGTDCTSIGTEIQSSSDQGIKRIKRAANPSKKUKIKIPQVLEAPGASKICGFCCHVAPQDSVYCSNDICILKHAATMKFLSSGKQKPKPKRMKQKPKSLPKCAQAGIKISVHRPAPKPKETTTVKAATVVVPARSEALGKEAACESSTPSWASRYNNAVKPKTAAPSPSLLYKCSGKLYLSLPSLIA"

1011..1160

misc_feature

/note="PHD; Region: PHD-finger. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains"

/db_xref="CDD:pfam00628"

ORIGIN

Alignment Scores:

Pred. No.: 4,036-111 Length: 2407

Score: 2811.00 Matches: 536

Percent Similarity: 96.43% Conservative: 5

Best Local Similarity: 95.54% Mismatches: 12

Query Match: 94.04% Indels: 8

DB: 9 Gaps: 2

US-09-787-016A-3 (1-562) x BC000770 (1-2407)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLeuLysProThrSer 20

Db 201 ATGACGACAAAGGCGACCCGAGCAATGAGGAGCGACCTAAGGCCATCAAAACCCAGC 260

QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrLeaLysArgGluGlyAla 40

Db 261 AANGAGTTCAGGAAACATGGGGTTTTCGAGGACCATATCGCCAAGCGAGGCGCGCA 320

QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60

Db 321 GGGGAGCGGGAGCGCTGACCCACTGAGGCGGCCACCCACAGCAGCAGCTGGGCGCTGTC 380

QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80

Db 381 CTGCGCGCGAGTGGGAGGAGCGGCCAAGCGCATCTGAGCGCGTGGAGCGTCTTGACCAT 440

QY 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100

Db 441 GGGCGGCGCGCGGCGAGGAGGAGCATGCTGTCTCCTCGAGGATTCGTGTCAGCCACG 500

QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120

Db 501 TCCTGCCCGCCACAGACGCCGAGCAGCTCCGAGGCGCAGCTGGAAGCGCTTCTGAG 560

QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140

Db 561 ACCAGAGCGCGCCCGAGTCTGTCTTCACAGCTGTGAAGGAACGACCGCTTCTTCTGNA 620

QY 141 LysValLysGlyLysAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160

Db 621 AAGGTGAAGAGGAGGAGGATGACCCAGCATGACCTCCGATAGTGAAGAGGATGGCTGACC 680

QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180

Db 681 TTGAAGAGCTTCAGATCGCTTCCAGAGAGCGGGAACAGAGCCACCTGAGAGGCGCC 740

QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGluProAlaGluThr 200

Db 741 CTGAAGAGGATCCAGATCGCTGCGGGAAGAGCGCGCGGAGGAGGGTCCCGCCGAGACT 800

QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220

Db 801 GTGGGCTCCGAGGCGCAGTGAAGTGTGGAGGCGGCTCTGCCAGTAAGCAGGAGGCCGAG 860

QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240

REFERENCE

AUTHORS

TITLE

JOURNAL MEDLINE PUBMED

REFERENCE AUTHORS JOURNAL

REMARK COMMENT

FEATURES source

gene

CDS

LOCUS MMU238332 2602 bp mRNA linear ROD 09-JUL-1999
Mus musculus mRNA for death inducer-obliator-1 (Dio-1).

DEFINITION AJ238332
ACCESSION AJ238332
VERSION AJ238332.1 GI:5457402
KEYWORDS death inducer-obliator-1; Dio-1 gene.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 Garcia-Domingo, D., Leonardo, E., Grandien, A., Martinez, P., Albar, J.P., Ipiñua-Belmonte, J.C. and Martinez-A.C. Dio-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development
Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999)

JOURNAL MEDLINE
PUBMED 99324176
REFERENCE 2 (bases 1 to 2602)
AUTHORS Garcia-Domingo, D.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology and Oncology, Centro Nacional de Biotecnología/CSIC, Campus Cantoblanco UAM, Ctra. de Colmenar Viejo Km 16, Madrid, 28049, SPAIN

FEATURES
source Location/Qualifiers
1..2602
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cell_line="WOL-1"
/cell_type="pre-BI"
1..2602
/gene="Dio-1"
232..>2073
/genes="Dio-1"
/function="apoptosis inducer"
/evidence="experimental"
/codon_start=1
/product="death inducer-obliator-1"
/protein_id="CA848401.1"
/db_xref="GI:5457403"
/db_xref="GOA:Q8C9B9"
/db_xref="GOA:Q9WV00"
/db_xref="SWISS-PROT:Q8C9B9"
/translation="MDKGLHNEAPKAIKPTSKERKTWGRFTTIKREGAGDTEADSEQQPQHNLRLSRSGPQKRTVERVEFLTVRRGKNVPSVLEDSSEPTSTVTDVETASGEVSSEIRSPVSDSLGKEHPASSEKAGGEEEDTSDSDGLTLKELOHLRRKEQEPVRSLSQNLKRRKGRKPECEVIDPDNALYICICROPHNN SQGVQSQSTDDIENQLEGKATNTENPREAGKPKPECEVIDPDNALYICICROPHNN RFMCCDCREWFHGCVGISEARGELLENEDGDIYCPNCTIQQVQDINGSATNEQD SGCRSGVAGDGTCTGTGTGKSGEDQGIKRIEKAANPSGKKGLKIPQVVEAGAP KCIVQVGIKISSVHKRLASEKRNPKVWKLASRSETSGKEAECSTPSWASHNTN AVXPEKPKTALSPITLLSKCTTHPKAGFPQPSHLGSCIGLSRTRVLGLVLIVASS SLFARSRYQASGQVFLPSLWLSLWGLKASCVGLMLEAISYFSFRP#"

gene
CDS

Alignment Scores:
Pred. No.: 1,94e-81 Length: 2602
Score: 2112.50 Matches: 422
Percent Similarity: 81.40% Conservative: 33
Best Local Similarity: 75.49% Mismatches: 94
Query Match: 70.68% Indels: 11
DB: 10 Gaps: 7

US-09-787-016A-3 (1-562) x MMU238332 (1-2602)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLysProThrSer 20

DB 232 ATGGATGATGATAAAGGGCCACCTGAGCAATGAGGAGGACCCAGGCTATCAACCCACCACT 291

QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40

RESULT 9
MMU238332

861 AACGATCAGGCGGTGTGTCCCGCTGGGAAGATCAGAGAGAGTAGTTGGAGGA 920

QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260

DB 921 AAGCGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGGCCACCGAGCTGAA 980

QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280

DB 981 TGTGAGGGTTACGACCCCAACGCCCTGTATTGCAATTTGGCCGCGCTCACACACAGG 1040

QY 281 PheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300

DB 1041 TTTATGATTGCTGTGACCGCTGTGAAGATGGTTTCATGCGATTGTGTGGCATTTCT 1100

QY 301 GluAlaArgGlyArgLeuLeuLeuArgAsnGlyGluAspTyrIleCysProAsnCysThr 320

DB 1101 GAGCGCTCGAGGAGGCTTTTGGAAAGGAATGGGAAGACTATATCTGCCCAACTGCACC 1160

QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyr 340

DB 1161 ATTCTGCAAGTGCAGGATGAGACTATTCAGAAACCGCAGATCAGCAGGAGCTAAATGG 1220

QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360

DB 1221 AGACCTGAGATGCTGTATGGCACCGAATTGTACAAGTATAGGAACAATAGACGAGTCT 1280

QY 361 SerGluAspGlnGlyIleLysGlyValGlyIleGluLysAlaAlaAsnProSerGlyLysLys 380

DB 1281 AGCGAAGCCAGGATTAAGGGTAGAATTGAGAGCTGCATTCAGTGGCAGAG 1340

QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400

DB 1341 AAACCTCAAGATCTTCCAGCTGTGTAGAGCGCGCTGCTGCTCAAAATGTTATGGCCCC 1400

QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLysLys 420

DB 1401 GGGTCTGTCTGCTGCGCGCAGCCGCTCGGTGTAATCGAGTAAGTACTGTATCTCAAA 1460

QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440

DB 1461 CACGCGCAGCGACCAATGAAGTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCTAAA 1520

QY 441 GlnLysMetLysMetLysProGluLysProSerIleuProLysCysGlyAlaGlnAlaGly 460

DB 1521 GAAAGATGAAGATGAAGCAGAGAGCCAGTCTTCGAAATGCGGTGCTCAGGACGT 1580

QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480

DB 1581 ATTAAATCTCTCTGTGCACAGAGACCAAGCTCCAGHAAAAAAGAGACCAAGTGAAG 1640

QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500

DB 1641 AAGGCAGTGTGCTGCTGCGCGAGTGAAGCACTCGGAGAGGAGGAGCTTGTGAGAGC 1700

QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520

DB 1701 AGCAGCGCTGTGGCGGAGCGCATCAATTAATCAATGCAATGCAATGCAATGCAATGCAAT 1760

QY 521 AlaProSerProSerLeuLeuTyrLysCys-----MetTyrHisLeuGlyVal 536

DB 1761 GCTCCCTCGCGCTGCTGTTGATAAATCTCTGTTGTAAGTATTGTTATTTCTTCTTCC 1820

QY 537 GlyLeuLeuAsp-----ProSerArgSerPheTrpIleAlaIleProTrpAla 552

DB 1821 AGTCTGATTGCTAGCCACACTGCCCGCGCGCGCCACATCCACCTCTGCTGCTGCTGCT 1880

-QY 553 Cys 553

DB 1881 TGT 1883


```
/mol_type="genomic DNA"
/db_xref="taxon:10095"

Alignment Scores:
Pred. No.: 2,14e-81 Length: 2867
Score: 2112.50 Matches: 422
Percent Similarity: 81.40% Conservative: 33
Best Local Similarity: 75.49% Mismatches: 94
Query Match: 70.68% Indels: 11
DB: 6 Gaps: 7

US-09-787-016A-3 (1-562) x BD244713 (1-2867)

Qy 1 MetAspAlaGluAlaAspProSerAsnGluGluAlaProLysAlaIleValProThrSer 20
Db 232 ATGGATGATAAGGCGACCTGACCAATGAGGAGCACCACCAAGGCTATCAACCCACCAAGT 291
Qy 21 LysGluPheArgGlyThrTyrGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
Db 292 AAGGAGTTTCAGGAAACCTGGGGTTCGAGACCAACGATGCGCAACGTCGAGGGTGA 351
Qy 41 GlyAspAlaGluAlaAspProLysGluProProProGlnGlnGlnLysGlyLeuSer 60
Db 352 GGAGACACGAGGCGGACCCACAGTCAGCAGCAACA-----CAGCAGCATAACCTCTCC 405
Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 406 CTGCCCGCAGTGGAGCGGACCAACACGCTACTGAGAGGAGTGAAGAGTTCTTACACAG 465
Qy 81 AlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 466 GTTCGGCGCGGAGGAGGAAAGAAATGCGCGGTGTCCTGGAGATTCCAGTGGAGCCACA 525
Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 526 TCTTCCACAGTCACATGATGAGACAGCTTTCGAGGGGAGCGTTGAAGCAGATTTCTGAG 585
Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 586 ATCAGAGTGGCCCTCTATCTGACCTCTAGGG---AAGAACATCTCTGCTCTCTGAA 642
Qy 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 643 AAGGCAAAAGGAGGTGAGAGGAGAGACACCTCTGACAGTGCAGAGTGAATGCGCTTAG 702
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGlnGlnGluProThrGluArgPro 180
Db 703 TTGAAGAACTTCAGAACCGCTTCGAGAAAGCGGAGCAGACCAAGACTCTGAGAGGTCC 762
Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGluProAlaGluThr 200
Db 763 CTGAGAGGCGAGTCAGAAATCGCTGAGAAAGAGCGCAGAGGAGAAATCTCTGCGGAACT 822
Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
Db 823 ---GGAGGTGTCCAATAGCAGTCGCGAGCAGGACAGACCTCTCTGTNAGCAGAGCT 879
Qy 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
Db 880 GAGGTAGTCAGGACCACTGTCCTCAGTCAGAGACAGATGACATAGAAATCAGTTGGAA 939
Qy 240 GlyLysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
Db 940 GGGAGGCGCAGTCAGGGAATCAGAGGAAACCCCGAGGAGCGGGCAACCAAGGCT 999
Qy 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
Db 1000 GAGTGTGAGTTTACGACCCCAATGCCCTGTACTGCTATCTGCGCGCAGCTCACAACAAC 1059
Qy 280 ArgPheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIle 299
Db 1060 AGGTTTATGATCTGCTGATCGGTGAGGAGTGGTTCCATGGTACTGTGTGGGTATT 1119

300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
1120 TCTGAGCCCGAGCGCGCTCTCGAAAGAAACGGGAGACACTACTCTCCCAATATGC 1179
320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
1180 ACCATTTTGAGTGTGAGGATGAGACAAACGCTAGCGCCCAATGAGCAGGACTCTGGG 1239
340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
1240 TGCAGATCTGTGGTGTCTGATGTCACAGACTGCACAGCAGTAGCGACAGTAGACAGG 1299
360 SerSerGluAspGlnGlyIleLysGlyValGlyIleLysAlaAlaAsnProSerGlyLys 379
1300 TCCGAGAGAGACCGGCGCATTAAGGCTAGGATGAGAGGCGCAGCAACCCCGCGCAAG 1359
380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
1360 AAAAACTCAAGATATTCAGCTCTGCTAGAGGCTCTCTGCTCTCTCTCTCTCTCTCTCT 1419
400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 419
1420 CCTGGGTGTTCAGGTGAGCAGCCCTGACTCTGTGTATGCGATATGACTGCTATCTC 1479
420 LysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
1480 AAACACCGCAGCAGCTACCATGAGATTTCTAAGTTTCAGTAAAGAAACAAAAACCC 1539
440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGluAla 459
1540 AAGAAAGGTCAAGACGAAAGCCAGAAAGTTTCAAGTTCTTCAAAATGCAAGTTCAGGTG 1599
460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVal 479
1600 GGGATTAATCTCTCTGTGTCACAGAGACTAGCGTCAGAGAAAGGAAAGCCAGTG 1659
480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
1660 AAGAAA---GTGATGTGGCTTCAGAGTGAAGTCTTCTGGAAGGAGCGAGCTGTGAG 1716
500 SerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThr 519
1717 AGCAGCACACCTCTCTGGCAGTGAAGTGAACCAACTACATGCTGTGAGCAGCAGAGGCCA 1776
520 AlaAlaPro-----SerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyVal 536
1777 GAGAAGCCCACTGCACTCTCGCCCACTATTGAGTAAATGATGATATCACCCCAAGGCT 1836
537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
1837 GGTTCACAGGCCCTCTCCATC--ATCTGGGTGGCT---GCTGGGGCTGTCTAGTA 1888

RESULT 11
AX023370 2867 bp DNA linear PAT 15-SEP-2000
LOCUS Sequence 2 from Patent WO0015787.
DEFINITION AX023370
ACCESSION AX023370.1 GI:10183782
VERSION
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 Leonardo, B., Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A.
Genes encoding for the human and murine death in
ducer-obliterators-1
Patent: WO 0015787-A 2 23-MAR-2000;
LEONARDO ESTHER (ES); MARTINEZ PEDRO (ES); CONSEJO SUPERIOR
INVESTIGACION (ES); DOMINGO DAVID GARCIA (ES); MARTINEZ ALONSO
CARLOS (ES); BANNERMAN DAVID GARDNER (GB); GRANDIEN ALF (SE)
Location/Qualifiers
```

```

source      1. .2867
            /organism="Mus sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:10095"

ORIGIN
Alignment Scores:
Pred. No.:      2,14e-81      Length:      2867
Score:          2112.50      Matches:      422
Percent Similarity: 81.40%      Conservative: 33
Best Local Similarity: 75.49%      Mismatches: 94
Query Match:    70.68%      Indels: 11
DB:             Gaps: 7

US-09-787-016A-3 (1-562) x AX023370 (1-2867)

QY      1 MetAspAspIyeglyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
DB      232 ATGCATGATAAAGGACACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCACCACT 291
QY      21 LysGluPheArgIySerThrTgGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
DB      292 AAGGAGTTACGGAACCACTGGGGTTTTTCGAAGAACCAACGATTCCTCAAGCGTAGGGTGCA 351
QY      41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
DB      352 GAGACACGAGGCGGACCCCTAGTACGACCAACCA-----CAGCAGCATTAACCTCTCC 405
QY      61 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB      406 CTCGCGCGCAGTGGAGCGCAACCAACCGTACTGAGAGGGTAGAAGGTTCTTACCAACG 465
QY      81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB      466 GTTCGCGCGCGAGGAAAGATGTGCGGTGTCTCCGAGGATTCAGTACGAGCCACCA 525
QY      101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB      526 TCTTCCACAGTCTCATGTGGAGACAGCTTCCGAGGGGCGGTGGAAGCAGCTCTCTGAG 585
QY      121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB      586 ATCAGAGTGGCCCTGTACTCTCTAGGG---AAAGAACATCTGCTCTCTGAA 642
QY      141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB      643 AAGGCAAAAGAGGTGAAGAGGAAGAAGACACCTCTGACAGTACACATGATGCGCTTACG 702
QY      161 LeuLysGluLeuGlnAsnArgLeuArgArgGluGlnGluProThrGluArgPro 180
DB      703 TTGAGGAACCTTCAGAACCCCTTCGAGAAAGCAGAGCAGACACTGTGGAGAGGTCC 762
QY      181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluGlyProAlaGluThr 200
DB      763 CTCGAGGCGAGTCAAGATCGCTGAGCAAGAGGCGCAGAGAGAGATTCTGCGGAAACT 822
QY      201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
DB      823 ---GGAGGTGTCGAATAGGAGTGGCGGACGACGACGACACTCTCTGTAGCAGGAGCCT 879
QY      220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
DB      880 GAGCGTAGTCAGGACCCAGTGTCCAGTCAGACAGACAGATCAGATAGAAAATCAGTTGGA 939
QY      240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
DB      940 GGGAGCGGACTCAGGGAATATACAGGAAACCCCGAGGAGCGGCGGCAACCAAGCCT 999
QY      260 GluCysGluGlyTyArgProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsn 279
DB      1000 GAGTGTGAGGTTTACAGCCCAATGCGCTGTATGTATCTGCTGCGCCGACCTGACACAC 1059
QY      280 ArgPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIle 299

```

RESULT 12

```

AK129117
LOCUS      AK129117              4906 bp      mRNA      linear      ROD 21-NOV-2003
DEFINITION Mus musculus mRNA for mKIAA0333 protein.
ACCESSION  AK129117
VERSION    AK129117.1  GI:37359897
KEYWORDS   FLI CDNA.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

REFERENCE
  1 Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S.,
    Suga, Y., Nagase, T., Ohara, O. and Koga, H.
    Prediction of the coding sequences of mouse homologues of KIAA
    gene: III. the complete nucleotide sequences of 500 mouse
    KIAA-homologous cDNAs identified by screening of terminal sequences
    of cDNA clones randomly sampled from size-fractionated libraries

```

JOURNAL DNA Res. 10 (4), 167-180 (2003)
 MDLINE 22977043
 PUBMED 14621295
 REFERENCE 2 (bases 1 to 4906)
 AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
 TITLE Direct Submissiion
 JOURNAL Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research
 Institute, Laboratory for Genome Informatics; 2-6-7
 Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan
 (E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
 COMMENT The CREATE program supported by Japan science and technology
 corporation; cDNA full insert sequencing; Kazusa DNA Research
 Institute; cDNA library construction, clone selection and 5'- &
 3'-end one pass sequencing.

FEATURES
 source
 1..4906
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="mbgl6015"
 /tissue_type="brain"
 /dev_stage="adult"
 /note="vector:modified pBC SK+"
 1..4906
 /gene="mKIAA0333"
 <224..3829
 /gene="mKIAA0333"
 /note="CDS is predicted by in silico analysis. Start codon
 is not identified."
 /codon_start=1
 /evidence="not experimental"
 /product="mKIAA0333 protein"
 /protein_id="BAC97927.1"
 /db_xref="GI:37359898"
 /translation="PVVFPKAGFOAFSGFGMDKGLHLSNEAPKAIKPTSKPRFKT
 WGPRTTIAKREGAGTEVDSEQQPHNLISLRSGQPKRTERVEFLTVRRCK
 KNVPSLSDSEPTSTVDTEASEGVSESSIRSGFVSDLSKHPASERKAKGG
 EEEEDTSDSDGLTKSLQNLRRKQEPVSRSLRGLRQNLRRKREEDSAETGSV
 QIGSAEQRLPKQEPFASQGVSDTDIENQLEKATQNTENPREAGKPKPEC
 EYVDNALYCIQCPHNRPMICDRCEWPHGDCVGISEARGLLENRGHDYICPNC
 TILQVDTGTSATDEQSGCRSCVADGTDCTSIGTVQKSGDQGIKGRLEKAAPS
 GKXKLIQPVVEAPGAPKICGPGCSVAQPDVYCSNDILKHAATWRLSSGKEQ
 KTKPEKVTXPKESLPCQSVQVGIKISSVHKLASREKPNVKKMLASRSTSGK
 EACSSSTPSWASDHNINAVKPEKPKTALSPILLSSKMDRDRVERDWTAAVTIPK
 KALPASLVGQTSRPNVFKLPPYSNMGAKPAIKLPGSGFKGTIPKRPVSATILS
 GTSARAGPTMTAAASKLPGSAVGVTRKPMSPNPAAPGLRGPVSPAPSQN
 SQIRQNIIRSLKILKRWNSDDLIWTENEVGKIALHIEKEMFLFQVTDNRYSKY
 RSIMFNLPKPNQGLFHRVREEISLAKLVKMKPEELVSKLSNMWTKPTKSVIESRT
 KLIANSKNTKPTETIPDMESPPVSDSEQESVRAAPEKSAAPLLDVFSSMLKDTT
 SCRAHLFDLACKICTGQVPSSEDEPAKKKLSASCKEDFKPHDSPPNVAVENTA
 DEGIADTLPENASEPDSSTSLNQRKCTFPESPGDSHPFESSLGLSPSASGSGSV
 VTTVTMSGRDPRALSGSCTVTASMAHLNDSQASSETKLDIMKPAITSAVVPKSLAK
 PSSSPDPRLSYPPSPSISESRSPPEGDTTLPLSLNTIWKGFIMQSVAKFVTKAYP
 VSGCLDLSLEDLPDTHIGGRAPKTVWDVYVGLKSSVSKLCIRFHPATSEEVAY
 ISLSYVFSRGRFGVANNRHRVKDLYLPLSAKDPVPSKLIIPFGPKGHPVGR"

ORIGIN

Alignment Scores:
 Pred. No.: 7,72e-80 Length: 4906
 Score: 2081.50 Matches: 412
 Percent Similarity: 82.80% Conservative: 31
 Best Local Similarity: 77.01% Mismatches: 83
 Query Match: 69.64% Indels: 9
 DB: 10 Gaps: 6

US-09-787-016a-3 (1-562) x AK129117 (1-4906)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 278 ATGGATGATAAGGGCACCTGAGCAATGAGGAGCACCCCAAGCTATCAACCCACCACT 337
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40

Db 338 AAGGAGTTTCAGGAAACCTCGGGTTTTTCGAAGAACCAAGATTGCAAAACGTGAGGGTGCA 397
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 398 GGGAGACACGGAGGTGGACCCAGTCAGTACCAACCA-----CAGCAGCATAAATCTTCC 451
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 452 CTGGCCCGCAGTGGACGGCAACCAACCTACTGAGAGGGTAGAAGAGTTTCTTACCACG 511
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 512 GTTCGGCGCCGAGGAGAAAGATGTGCGCGTGTCCCTGGAGGATTCCAGTGAGCCACACA 571
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 572 TCTTCCACAGCTCTGATGTGGACAGCTTCGAGGGGGAGCGTTGAAAGCAGTTCTGAG 631
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 632 ATCAGAAGTGGCCCTGTATCTGACTCTTAGGG---AAAGAACATCTCGCTCTTCTGAA 688
 QY 141 LysValLysGlyArgAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 689 AAGCAAAAGAGGTGAAGAGGAAAGAACACCTCTCAGTGCAGTGTGATGGCTTACG 748
 QY 161 LeuLysGluLeuGlnAanArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
 Db 749 TTGAAGCAACTTCAGAACCGCTTCGAGAAAGCGAGAGCAAGAACCTGTGAGAGGTCC 808
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
 Db 809 CTGAGGGCAGTCAAGATCCGCTCAGGAAGAGCGAGAGGAAGATTCTGCCGAACT 868
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 Db 869 ---GGAGTGTCCAAATAGGCAGTGGCCGAGCAGCAGACCTCTCTGTAGCAGAGCCT 925
 QY 220 GluAsnAspGlnGlyValValSerGluAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 Db 926 GAGCTAGTCAGGGACCAAGTGTCCAGTCAGACAGATGATGATAAAATCAGTTGTGAA 985
 QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyLysArgLeuGlyArgProLysPro 259
 Db 986 GGAAGGCGACTCAGGANAATACAGAGAAACCCAGGAGGAGCGGCAACCAAGCCT 1045
 QY 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
 Db 1046 GAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGTCCATCTGCGCCAGCCTCAACAAC 1105
 QY 280 ArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIle 299
 Db 1106 AGGTTTATGATCTGCTGTGATCGGTGTGAGAGGTGGTTCCATGTGACTGTGTGGTATT 1165
 QY 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
 Db 1166 TCTGAGGCCCGAGCGCGCTCTCGAAGGAACCGGGAAGACTACATCTGCCCAATTCG 1225
 QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLys 339
 Db 1226 ACCATTTCGAAGTGCAGGATGAGCAACCGTAGCGCCACCGATGACAGACTCTCGG 1285
 QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
 Db 1286 TGCAGATCTGTGGGTGTGTGTCGACAGACTGCACAGCAGTAGGACAGTAGAGCAGAAG 1345
 QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGlyLysAlaAlaAsnProSerGlyLys 379
 Db 1346 TCCGAGAGAGCCAGGCGATTAAGGGTAGGATTGAGAGGAGGAGCAACCCCGCGGCAAG 1405
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 Db 1406 AAAAATCTCAAGATATTCCAGCTGTCTGTAGAGGCTCTGTGTCTCTTAATGATTCGTC 1465


```

QY 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 419
Db 1466 CTTGGTGTTCAGTCTAGCACAGCTGACTCTGTGTATTGAGTAATGACTGCAATTC 1525
QY 420 LysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
Db 1526 AACACGCGAGCTACCATGAGATTCTAAGTTTCAGGTAAAGAACAAAAACAAACCC 1585
QY 440 LysGluLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
Db 1586 AAGGARAAGTCCAGACGAGCAGAAAGATTCCAGTCTCCAAATGCAGTGTCCAGGTG 1645
QY 460 GlyIleLysSerValHisLysArgProAlaProGluLysGluThrVal 479
Db 1646 GCGATTAAATCTCTCTGCAAGAGACTAGCGTCAGAGAAAGGAGAAACCCAGTG 1705
QY 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGlu 499
Db 1706 AAGAAA---GTGATGTGGCTTCAGAGGTGAGACTTCGGGAAGGAGGAGCTGTGAG 1762
QY 500 SerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThr 519
Db 1763 AGCAGCACCACTCTCTGGGAGTGCACCACTACAAATCTGTGAAGCCAGAGAGCCCA 1822
QY 520 AlaAlaPro-----SerProSerLeuLeuTyrLysCysMet 531
Db 1823 GAGAGCCCACTGCTCTGCCACCTTCCCTATTCAGTAATCCATG 1867

RESULT 13
AB002331
LOCUS Human mRNA for KIAA0333 gene, partial cds. 6692 bp mRNA linear PRI 06-OCT-2001
DEFINITION Human mRNA for KIAA0333 gene, partial cds.
ACCESSION AB002331
VERSION AB002331.1 GI:2224606
KEYWORDS KIAA0333.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Nagase, T., Ishikawa, K., Nakajima, D., Chira, M., Seki, N.,
Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
Prediction of the coding sequences of unidentified human genes.
VII. The complete sequences of 100 new cDNA clones from brain which
can code for large proteins in vitro
JOURNAL DNA Res. 4 (2), 141-150 (1997)
MEDLINE 9434984
PUBMED 9205841
REFERENCE
2 (bases 1 to 6692)
Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.
Direct Submission
Submitted (28-MAR-1997) Osamu Ohara, Kazuo DNA Research Institute;
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)
Location/Qualifiers
1 .6692
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG0988"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1 .6692
/gene="KIAA0333"
<1 .2978
/gene="KIAA0333"
/codon_start=3
/protein_id="BA020791.1"
/db_xref="GI:2224607"
/translation="ESKLEKAAQDINDSEPGDLGPKPECEGYDNALYICIRQPHN
NRPMICDRCEEFHGDGVGISBARGILLERNGEDYICFNCTILQVDETHSETADQQ

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3,35e-58 Length: 6692
Score: 1575.00 Matches: 295
Percent Similarity: 89.13% Conservative: 0
Best Local Similarity: 89.13% Mismatches: 0
Query Match: 52.69% Indels: 36
DB: 9 Gaps: 1

US-09-787-016A-3 (1-562) x AB002331 (1-6692)
QY 235 GluSerLysLeuGluGlyLysAlaAlaGlnAspLysLeuAspGluGluProGlyAspLeu 254
Db 3 GAGAGTAAGTTGGAGGAAAGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTG 62
QY 255 GlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArg 274
Db 63 GCGCCGACGAGCGCTGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCATTTGCCGC 122
QY 275 GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluThrPheHisGly 294
Db 123 CAGCTTCACAAACAGAGTTTATGATTCTGTGACCGCTGTGAAGATGGTTTTCAGGC 182
QY 295 AspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr 314
Db 183 GATTGTGGGCAATTCCTGAGCTCGAGGAGGAGCTTTTGAAAGGAATGGGAGACTAT 242
QY 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
Db 243 ATCTGCCAAACTGCACCAATCTGCAAGTGCAGATGAGACTCATTCAGAAAACGGCAGAT 302
QY 335 GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
Db 303 CAGCGGAAGCTTAAATGGAGACTGGAGATGCTGATGGCAGCGATTTGTACAGTAGGA 362
QY 355 ThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAla 374
Db 363 ACAATAGAGCAGAGTCTAGCGAAGACCAAGGAGTAAGAGGATGAGATTGAGAAGCTGCA 422
QY 375 AsnProSerGlyLysLysLeuLysIlePheGlnPro----- 387
Db 423 AATCCAAAGTGGCAGAGAAACTCAAGATCTTCCAGCTGTGTCGCCGGCTGTGCCACC 482
QY 387 ----- 387
Db 483 CAGTCCTCTCTCTGCGAGGATTTTGAAATGTCTGTCTAGAGCATCTCTGCTTTC 542
QY 388 -----ValIleGluAlaProGlyValaSerLysCysIle 398
Db 543 ACACCTTTCACATGCAATTAGCTGCAAGTGAATAGAGCGCTGTGCTCCATCATGATATT 602
QY 399 GlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIle 418
Db 603 GCGCCCGGTGTGTGTACGTGGCGAGCCCGACTGGTGTACTGCAAGTATGACTGTATC 662
QY 419 LeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 438

```

DB	663	CTCAACACGCGCAGCGACCAATGAAAGTTCTTAGCTCAGGTAAAGACAGACGCCAAAG	722
QY	439	ProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGln	458
DB	723	CCTAAAGCAAGATGACATGACGAGCGAGAGCGCGCTTCGGAATGCGTGCTCAG	782
QY	459	AlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThr	478
DB	783	GCAGGTATTAAATCTCTTCTGTGCGACAGAGACCGAGCTCCAGAAAAAAGAGACACCA	842
QY	479	ValLysLysAlaValValProAlaAlaArgSerGluAlaLeuGluLysGluAlaAlaCys	498
DB	843	GTGAAGAGCGAGTGTGTGCTCCCTGCGGAGTGAAGCACTCGGAGGAGGAGAGCTGTGT	902
QY	499	GluSerSerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLys	518
DB	903	GAGAGCAGCAGCGCGTGTGCGGAGCGATCACAATTAACATGACGTAAAGCCAGAAAG	962
QY	519	ThrAlaAlaProSerProSerLeuLeuTyrLys	529
DB	963	ACTGCTGCTCCCTGCGCGCTCACTGTTGTATAAA	995
RESULT 14	AC108337	272404 bp DNA linear HTG 21-SEP-2002	
LOCUS	Rattus norvegicus clone CH230-115113, *** SEQUENCING IN PROGRESS		
DEFINITION	***, 2 unordered pieces.		
ACCESSION	AC108337		
VERSION	AC108337.4 GI:23266147		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 272404) Muzny, D. Marie., Metzker, M. Lee., Abruzzo, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benatmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebowitz, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, P., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milobavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,		

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soja, J., Steinfle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, F., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Welles, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Unpublished
Direct Submission
2 (bases 1 to 272404)
Worley, K. C.
Direct Submission
Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 272404)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21737648.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPWX
Center clone name: CH230-115113
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 238513 bases at least Q40
Consensus quality: 241910 bases at least Q30
Consensus quality: 244206 bases at least Q20
Estimated insert size: 262154; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
* NOTE: This sequence may represent more than one 'clone'-
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 270622: contig of 270622 bp in length
* 270623 270722: gap of unknown length
* 270723 272404: contig of 1682 bp in length.
Location/Qualifiers
1. 272404
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-115113"
1. 1256
/note="wgs_end_extension"

FEATURES
source
misc_feature
1. 1256

misc_feature
clone end:Sp6"
complement(1814..2663)
/note="clone boundary
clone end:Sp6
site:EcoRI
end sequence:BH294125"
62809..64046
/note="wgs contig"
267959..268006
/note="clone boundary
clone end:r7
site:EcoRI
end sequence:BH294091"

ORIGIN

Alignment Scores:
Pred. No.: 6.06e-48 Length: 272404
Score: 1372.50 Matches: 312
Percent Similarity: 44.34% Conservative: 25
Best Local Similarity: 41.05% Mismatches: 55
Query Match: 45.92% Indels: 368
DB: 2 Gaps: 7

US-09-787-016A-3 (1-562) x AC108337 (1-272404)

Qy 1 MetAspAspGlyAspProSerAsnGluGluAlaProlysAlaIleLysProThrSer 20
Db 128283 ATGATGATTAAGGGGAGCTCGAGCAATGAGGAGCACCAGGCTATCAAAACCACCAGT 128342
Qy 21 LysGluPheArgGlyThrTgGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 128343 AAGAGTTTCAGAAACCTGGGGTTTTTCAGAGAACCAATTCGCAACGAGGGTGCA 128402
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
Db 128403 GGACACCGAGGTGGAGCCCGAGTGAGCAGCACCA-----CAACAGCATAACTCTCC 128456
Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 128457 CTGCGCGCAGTGGACGCGACCCAAAGCGGACCGAAAGGGTAGAAGAGTTCCTTACTACT 128516
Qy 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 128517 GTTCCGCGCGAGGAGGAGAAATGTCGCGTGTCCCTGGAGGATTCAGTGAGCCCAAG 128576
Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 128577 TCTTCCACAGTCACTGATGTGGACACAGCTTCTGAGGGAGCGTTGAAGCAGTCTTGAG 128636
Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 128637 ATCAGAAAGTGCCTT---GCACCTGGCACCTTAGGAAAGACATCTTCTCTTCTGAA 128693
Qy 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 128694 AAGGTGAAGAGGCTGATGAGAGAGAGACACCTCTGACAGTGACAGTGATGCTTACG 128753
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Db 128754 CTGAAGGAACCTTCAGAACCGCTTCGAGAAAGCGAGCAGGAGAACCTGTGGAGAGTCC 128813
Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
Db 128814 CTGAGAGGCATTCAAGATCGACTCGGAGCAGCGCCGAGAGGAGATCTCTGCCGAAGCT 128873
Qy 201 -----ValGlySerGluAlaSerThrValGluGlyValLeuProSerLys 216
Db 128874 GGGAGTGTCCAGATAGGAGGATTTGAGGAGGACACACCTCGC-----TGTAAG 128921
Qy 217 GlnGluProGlnAsnAspGlnGlnValValSerGlnAlaGlyLysAspAspArgGluSer 236
Db 128922 CAGGAGCTGAGGCTAGCCAGGGACCGAGTGTCCCGAGTCAGAGAGAGATGACAGAGAAAT 128981

Qy 237 LysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArg 256
Db 128982 CAGTTGGGGGAGAGCGACTCAGGGAAGTAAAGGAGAAAGCCCGAGGAGCTGGCAA 129041
Qy 257 ProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnPro 276
Db 129042 CCAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCCGCAGCT 129101
Qy 277 HisAsnAsn----- 279
Db 129102 CACACAAACAGGTAGTGGAGGGAGGGAAGTGTAGAGAACCAAGGATGGAGACAGCGA 129161
Qy 279 ----- 279
Db 129162 GAGGACTGACATTGCTTTTATGTCAAAAGCAGTGTGGCTTATTTGAGAAAGAGCAATGA 129221
Qy 279 ----- 279
Db 129222 GTCTGAAAAAAGAAACAGAGATGCAATGATGATGCACTGCCCTAGAGCTGCCCTCTACCATG 129281
Qy 279 ----- 279
Db 129282 GCACCCACGTTGCTCCATGTCAGGGGATGAATGACACAGTGCCTCAGTCTCTCAAGCCAT 129341
Qy 279 ----- 279
Db 129342 GTCCAGATGGACACTGTAGATAGCGGAGAACCAACTGTTGTCTCTACAGATCTTATGGGAG 129401
Qy 279 ----- 279
Db 129402 CGTTAGAGGATCCACCCCTGTATGTCCTTTGAGTGTGACTGCCAGTGTGTGATGGGGCATC 129461
Qy 279 ----- 279
Db 129462 TCTGTGGATTTGTAGAGGTGAAGTTTCTGTTGCTGTGTGGAGATACTCTAAAGCCTCT 129521
Qy 279 ----- 279
Db 129522 AATTCACTCTTAGTGTCTTAACCCAGAGATGAAGGGCTGAGGATGTGGCTCTGTGAGTAG 129581
Qy 279 ----- 279
Db 129582 AATGCTTACTCAGCACACATGAAGCCCTGGGTGTGATGCCAGTACCTCATATAATGGT 129641
Qy 279 ----- 279
Db 129642 AGTGGCATTATGCTGTAACTCAGCTCTCAGTGTAGAGGAGGAGATCAAGAAATTCAGA 129701
Qy 279 ----- 279
Db 129702 TCATCTTTGGCTAGCTAGGCGAGTTTAAGACAGGAGAAACCCCTGTCTCTAAAAGTAAGTAA 129761
Qy 279 ----- 279
Db 129762 ATAGAGAAAAAGATGAGGGGCTGGAGAGATGCTCGGTGGTTAAGAGCACTGACTGCTC 129821
Qy 279 ----- 279
Db 129822 TTCCAGAGGTCTGAGTTGATTTCCAGCAACCAATGGTGGCTCAACACCATCTGTAAT 129881
Qy 279 ----- 279
Db 129882 GAGATGCGARGTCTCTCTTCTGTGTGTCTGAGACAGCCTACAGTGTACTCATTAATTA 129941
Qy 279 ----- 279
Db 129942 TTATTAAATAAGTAATAACAGTGTATACATTAAGTCTTTTTTAAATAAGAAAAAGAT 130001
Qy 279 ----- 279
Db 130002 GAAGCAGATTGGAGTGTCTTAACAAAGATGGGGCACTTGGATGTTTTTAGCTCTG 130061
Qy 279 ----- 279

Db 130062 ATCTCACTTTCCCTTTATTTGCGATCTCTTCACTACTAGATGTTGAAGTGTGACCCACA 130121
 Qy 280 -----ArgPheMetIleCysCysAspAr 287
 Db 130122 TCATCTGTGGACAGTACACCATGATCTTCCATTTAGTTATGATCTCTGTGACCG 130181
 Qy 287 GCYSGlulThrPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLe 307
 Db 130182 GTGTGAAGTGTTTCCATGGGATGCGTGGGATATCTGAGGCCCGGAGGCGGCTCTCT 130241
 Qy 307 uGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspG 327
 Db 130242 GAGAGGAATGGGAGGACTACATCTGCCAAATTTGCCAAATTTTGCAGTGCAGGATGA 130301
 Qy 327 uThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyrPargProGlyAspAlaAspG 347
 Db 130302 GACGAGTAGTACGCGCGCGATGAGCAGGATGCTGGGTGCGAGACCTGTAGTGTCTGARGG 130361
 Qy 347 YThrAspCysThrSerIleGlyThrIleGlnLysSerSerGluAspGlnGlyIleLY 367
 Db 130362 CACAGACTGTACACCATAGGACAGTACAGAGTCCCGAGAGACGAGGCGCATAAA 130421
 Qy 367 sGlyArgIleGluLysAlaAlaAsnProSerGlyLysIleLysIlePheGlnPr 387
 Db 130422 CGGTAGGATTGAGAGGCGGCAACCCCGAGCGGCAAGAAAGCTCAAGATATTTCCAGCC 130481
 Qy 387 oValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAla 406
 Db 130482 CGTA-----AGTACCTGCTG---CCTTGGCAGGCGACATGTGTCA 130518
 RESULT 15
 AL732560/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP24-209P15 on chromosome 2, complete sequence.
 ACCESSION AL732560
 VERSION AL732560.11 GI:22474418
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 104663)
 Leongamornlert,D.
 Direct Submission
 Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Aug 23, 2002 this sequence version replaced gi:22204532.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-209P15 is from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

FEATURES

source
 1. 104663
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP24-209P15"
 /clone_lib="RPCI-24"

ORIGIN

Alignment Scores:
 Pred. No.: 4,44e-47 Length: 104663
 Score: 1342.00 Matches: 422
 Percent Similarity: 22.93% Conservative: 32
 Best Local Similarity: 21.31% Mismatches: 95
 Query Match: 44.90% Indels: 1433
 DB: 10 Gaps: 10
 US-09-787-016a-3 (1-562) x AL732560 (1-104663)
 Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 51952 ATGATGATTAAGGACCTGACCATGAGAGACCAAGGCTATCAACCCACCCAGT 51993
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 Db 51892 AAGGAGTTCCAGAAAACTGGGGTTTTTCGAAGAACCACGATTCGCAACCGTAGGGTCA 51833
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 51832 GGAGACACGAGGTGGACCCCGACGTGACGACCAACCA-----CAGCAGCATAACTCTCC 51779
 Qy 61 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 51778 CTGCGCGCAGTGCAGCGCAACCAACCAACCTACTGAGAGGGTAGAAGAGTTTCTTACCACG 51719
 Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 51718 GTTCGGCGCGAGGAAAAAGATGTCGGGTGTCCTCGAGGAGTTCAGTGACCCACA 51659
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGlySerValGluSerAlaSerGlu 120
 Db 51658 TCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGAGCGTGTAAAGCAGTCTTGAG 51599
 Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 51598 ATCAGAAAGTGGCCCTGTATCTGACTCTCTTAGGG---AAAGAACATCTCGCCTCTTCTGAA 51542
 Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 51541 AAGCAAAAGAGGTGAAGAGGAAGAAGACACTTGCACGTGACGTGATGGCTTAAG 51482
 Qy 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
 Db 51481 TTGAAGGAACTTCAGAACCGCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGGTCC 51422
 Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200
 Db 51421 CTGAGAGCGAGTCAAGATCGGCTGAGGAGAGCGCAGAGAGGAAGATTCTGCCAACT 51362
 Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 Db 51361 ---GGGAGTGTCCAAATAGGCAGTGGCGAGCAGGACAGACCTCTCTGTAGCAGAGCGCT 51305
 Qy 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 Db 51304 GAGCTAGTACGGGACCGAGTGTCCAGTCAGACAGACAGATGACATAGAAATTCAGTTGGA 51245
 Qy 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyLysArgProLysPro 259
 Db 51244 GGGAGAGCGCATCTCAGGCGAATACAGAGGAAAAACCCAGGGAGGCGGCAACCAAGCGCT 51185

QY 260 GluCySGluGlyTyrAspProbsnAlaLeuTyrCysIleCysArgGlnProHisLenAmn 279
Db 51184 GAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGTCATCTCGCCGAGCCTCACACAC 51125
QY 279 ----- 279
Db 51124 AGGTAGGTGGAGGAGGAGCTGGTAGAGAACCGAGGGGATGAGACAGTGGAGGAGCTGC 51065
QY 279 ----- 279
Db 51064 CATTTGCTTTGATGCCAAACAGCAGTTCGCTTTATTAGAAAAGATGAATGAAGCTGAAA 51005
QY 279 ----- 279
Db 51004 AAAGAAACAAAGAGAGAGCATGCACTGAGCTGCCCTAGAGTGCCTCTTACCATGGCAGC 50945
QY 279 ----- 279
Db 50944 CAGCTGTGTCAGTTCCTATACACTCAGTGTGTCCAGCCGGGATGAAGAGCTCACTGCCCT 50885
QY 279 ----- 279
Db 50884 CAGCTCTCAAGCCATGTGCAACCGTGGACACTGTGTAGATAGGCAAGAACCATCTGTTGT 50825
QY 279 ----- 279
Db 50824 CCCACAGATCCTATGAGAGGTGTAGAGGACATTAACAACACTGCTGTCTGCTTTGAAT 50765
QY 279 ----- 279
Db 50764 GGATGCCAGTGTGATGGGGCATCTCTGTGGGGTTGTTAGGAGATGAAGTTGCTGCAGC 50705
QY 279 ----- 279
Db 50704 TATTGTGAGTGAGATACCTCTGAAGCTCTAGTTCATCTTTAAAGTGTCTTAGCCAAA 50645
QY 279 ----- 279
Db 50644 AGATGAAGACTAAGGATGTTGTTTCATTGGGTAGAAATGCTTGCTTAGCATGCATGAAGCC 50585
QY 279 ----- 279
Db 50584 CTGGGTGTGATCCCGAGTATCTCATAAAGTTGTGAGGTGTATATGCTTATAATCAGCTT 50525
QY 279 ----- 279
Db 50524 TTAGTAGGGCAAGAGAAATCAGAAATTCAGATCATCTTTGGCTACGTAGGAGTTTAAG 50465
QY 279 ----- 279
Db 50464 ACATGAAAAACCTGTCTCTAAAAGTAAATAGAGAAAAAGATGGGTGGACAGATAG 50405
QY 279 ----- 279
Db 50404 CTCAGCAGTTAAGAGCATGACCGCTGTTCTTCCAGGGGTCCTGAGTTCAGTGGCCAGAA 50345
QY 279 ----- 279
Db 50344 ACTAGGTGGTTGCTTAAACCATCAGTACCGGATCCATTTGCCCTCTTCTGTGTGTCTA 50285
QY 279 ----- 279
Db 50284 CAGACAGTACAGTGTGCTCANGTATATATAAAGTGAAGCAGATTTGGAGTGTGCTGCTAA 50225
QY 279 ----- 279
Db 50224 CAGGGCAGCTTGGGTGTTTTTAGTTCGATCTCACTTTTCACTTTTATTGCTGTGCTGTT 50165
QY 279 ----- 279
Db 50164 ACAATAGATGTTAAAGTTTGAATGCTGTGATCTATTGGACAGTGCACATATCTTCC 50105

QY 280 ----- ArgPheMetIleCysCysAsnArgCysGluGluTyrPheHisGlyAspCysVal 297
Db 50104 CATTTTAGGTTATGATCTGCTGTGATCGGTGTGAGGAGTGTTCCTAGGTGACTGTGTG 50045
QY 298 GlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysPro 317
Db 50044 GGTATTTCTGAGGCCCGAGGGGGCTCTCGAAAGAAACGGGAAGACTATACATCTGCCA 49985
QY 318 AsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlu 337
Db 49984 AATTGCACCATTTTTCAGTGCAGATGAGACAAACGGTAGCGCCACCGATGAGCAGGAC 49925
QY 338 AlalysTyrAlaArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlu 357
Db 49924 TCTGGGTGCAGATCTGTGGTGTCTGATGCGCACAGACTCCACAAGCATAGGACAGTAGAG 49865
QY 358 GlnIysSerSerGluAspGlnGlyIleIysGlyArgIleGluIysAlaAlaAsnProSer 377
Db 49864 CAGAACTCCGAGAGAGACACAGGGCATAAAGGGTAGGATTGAGAGCGCAGCAACCCGAC 49805
QY 378 GlyIysIysIysLeuIysIlePheGlnPro ----- 387
Db 49804 GGCAAGAAAACACTCAGATATTCAGGCTGTGAGTATCTGTCTGCCATGGCAAGCACAC 49745
QY 387 ----- 387
Db 49744 ARGTAGGCTTTCTGCTGTGAAGAGAGCAATACTAAGAAGGTGCTTCTCTTAGGACT 49685
QY 387 ----- 387
Db 49684 GTTGGGTGGGGCTTTTGTGGGCAGTTACTGTATCTGTACGGGCCCTGGGCCCTGCAAGAGA 49625
QY 387 ----- 387
Db 49624 ACAFTCTGCTGTGTACACACTTGGCACTCCACTTGGTGTCAATTATGACTGTCTCATCA 49565
QY 387 ----- 387
Db 49564 TAGCCTCTTAGTGCCAGCAGCTTACTGTGCTCTAATGTCTACTTAAGGGAAATCCATT 49505
QY 387 ----- 387
Db 49504 TCTCCACAGTGTCTATGTACTGAGTGTGAGAAAAAACCTTACTTTTTCACACGCTCTTG 49445
QY 387 ----- 387
Db 49444 GTGTTACATATTTTGTGAGTCTGGCAGAGTTATTGAGAGAAAGGTGCACTGTGCTCTGCTC 49385
QY 387 ----- 387
Db 49384 TTACTTACATGCTGGCCTAAAGTTGTTCATGCTGTACTCANGTGTTCATGTTTCTTCTTCC 49325
QY 387 ----- 387
Db 49324 TACTTGAGTTATGTTATCCATGGGGCCATAACTGTCAAAATTTCCCATGAGCTGAGACTA 49265
QY 387 ----- 387
Db 49264 GCAGTCTTCTGCTCAGGAGAGCTTTCTCTGCTGTCAACAGGTGTGTGCTATTCTCAG 49205
QY 387 ----- 387
Db 49204 GAGGGTGCCACCAAGTGCATCTTCACTTTTAAACCCCAACCCCAAGCACTCCAGCAGG 49145
QY 387 ----- 387
Db 49144 TGAGAGTCTGGGAAACTACAATCTAGACTGGCACTGTCTTGTAGTCTCTGCTGTGCTGCAA 49085
QY 387 ----- 387
Db 49084 GGTGAAAAAGTGAATGACATCAGATCACCGTGGCCACTAGCACCGTTTACTGAGGAGATT 49025
QY 387 ----- 387

Db 49024 GTCATTTGATGCTTTTATTATTCCAGTCAGTATGTAGACATTGCTGAATTTTAGTC 48965
Qy 387 ----- 387
Db 48964 TGTGTAACCTTAGATTGGTAGAAGAAAGCAACTATCTTAAATATTTATCTCGAAACT 48905
Qy 387 ----- 387
Db 48904 GGGAGCTCTCAAAATGTGCAGACATATGTTCTGTTCAGTCTTCTCTGTATATATCTAGACTA 48845
Qy 387 ----- 387
Db 48844 AACGGTCTCTAGTTTCTGAAATGATATCTCTGTCTCTTTTGGCAATCTCCCTCCAAA 48785
Qy 387 ----- 387
Db 48784 CCACTCAAAAACAACAAACAAAAACAGGCTGCTTTCTTCAAAACTAAACAGTGGATCTC 48725
Qy 387 ----- 387
Db 48724 TTCTCTCTTAGGGCAGGCACATTTACTCTCTTTAAGAAATTTCACTTGAAGGCTAGGGT 48665
Qy 387 ----- 387
Db 48664 GGTGGACTCCAACCTCTAAAGGCAGTATGATTCATATATATATATTTCCCTCTGAGGGC 48605
Qy 387 ----- 387
Db 48604 TGGAGAGCTGCTCAGATTCAGAGGAAGTACTAGTTCAATTCATAGTGGCCATATGGTTCTC 48545
Qy 387 ----- 387
Db 48544 ATCGTCATTTATAGTTCAAGTGTGAGGAGCTGTGATATCTCTGTGATGACAGACAGA 48485
Qy 387 ----- 387
Db 48484 CATGAAGCAAGACACACATACATATAAGGAATTCOAATTTTGTGACAGTTATTAAGAG 48425
Qy 387 ----- 387
Db 48424 GAATTTTACCTGAGCTCATAGATGGCTCATTCGATAATAGCATGGCCATCAAAATCTG 48365
Qy 387 ----- 387
Db 48364 ATTACCTGAGTTTAATCTTCAGATCTCAGTGGTAAAGGAGAACTGACTCTCTGAAAG 48305
Qy 387 ----- 387
Db 48304 TTGTCCTGACTCTGCTGCTACCGCTATTCATGTGCAAGCTGCGGTATATGCAATGC 48245
Qy 387 ----- 387
Db 48244 ATGCACACAAATATGTAACAGAAATCTTAATGAAATTTGTAACTCATTAAGTCTACAT 48185
Qy 387 ----- 387
Db 48184 TTTTGGCTCTTTTTCATAGTGGTTTATTTTAAATGTTGTCTAGGGGAGTAGT 48125
Qy 387 ----- 387
Db 48124 TGCAACAGTATAGATGCCCTCAGAGGACAGGGCTATCTGACATGAGTTCTAAACACTAGA 48065
Qy 387 ----- 387
Db 48064 CTGCTCTTCTGTAAGCATCCTCTGAACCATGTGTTTCAGCTCTCTCTTACTCTTACCT 48005
Qy 387 ----- 387
Db 48004 CATTTGGACCTGCACATGTTTAGAGGCAATAGATCTGAAGCTCCCATGTGTGCAGAG 47945
Qy 387 ----- 387

Db 47944 TTTGTGTGCCAGGCGAGTCAAGGAAATAGGTTTGTATAGTTTCACAAAGATAGGTTCTTAC 47885
Qy 387 ----- 387
Db 47884 CATGTAGCTGGGCTGACCTGAAATTTGGAAGTCTCTCGCTCCCTTCCCTCCATGTGCTAG 47825
Qy 387 ----- 387
Db 47824 AATGGTGTAAAGATTATGTATCATGTCACAGCTAACAAATGAAAAATGAATTTTATT 47765
Qy 387 ----- 387
Db 47764 CATATAAACTATTATTAGAGCCATCCATAGTATGTATACACTTGCACCTTTTAGTCCCAG 47705
Qy 387 ----- 387
Db 47704 CACTGAGGAGGAGAGCAGGCAAAATCTCTTGAGTTCAAGGCTGGCTGTGTCATATC 47645
Qy 387 ----- 387
Db 47644 AAGTTTCTGGCAGCCAGAAATTTCATAGTGAGACCCTCTCTCAAAAAACACCAATAAA 47585
Qy 387 ----- 387
Db 47584 AATTCCTGTAGTGTCTTAATACCTGAGCCAAATAGTGTGTCTCACTGAGCACTGATC 47525
Qy 388 -----Vallelglua 391
Db 47524 ACTAGTGTGACACATGATGTCTGAAGTTGCTGTCTGTCTGCGAGTCTGTAGAGGC 47465
Qy 391 aProGlyAlaSerLysCysIleGlyProGlyCysCysHisValaladlaProaspSerVa 411
Db 47464 TCTGTGTCTCTAAATGATTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 47405
Qy 411 lTyCysSerAsnAspCysIleLeuLysHisAlaAlaThrMetLysPheLeuSerSe 431
Db 47404 GTATTGCGATTAATGACTGCTCAACACAGCAGCAGCTACCATGAGATTTCTTAAGTTC 47345
Qy 431 rGlyLysGluGlnLysProLysProlLysMetLysMetLysProGluLysProSe 451
Db 47344 AGGTAAGAACAAAAACAAACCCCAAGGAAAGGTCAAGACGAGCAGGCGGCGGCGGCGGCGG 47285
Qy 451 rLeuProLysCysGlyAla----- 457
Db 47284 TCTTCCAAATGCGAG-TGTTTCAGTAAGTGTGACACAGTCACTCCTCAGGATACTTACTTG 47226
Qy 457 ----- 457
Db 47225 AAACAGGAAGTGTGTGAGAGGACTCTTTCATAGACTGGTGTCTTGTGTGTGAGCAATGC 47166
Qy 457 ----- 457
Db 47165 AGGGACAGCAGCTGTCTGAGAAATATAGTATATTTTCTCTCTCTCTCTCTCTCTCTCT 47106
Qy 457 ----- 457
Db 47105 CTGAGGATTTATAGTTAAATGATGATATATCTACAGCAAGCAACCATGTGGCTTTTTTAA 47046
Qy 457 ----- 457
Db 47045 AGAAGTATGACTTCAGGATGTATACACACAGCTTTATGCAAAAGTGCATCTGACATTTAG 46986
Qy 457 ----- 457
Db 46985 ACAGAGCATAGGTTCTTAGGCAAGGCTCTGTGCTTGTGATAAGTCAGTGATAGAGTGCCT 46926
Qy 457 ----- 457
Db 46925 AGCATGCAAAACAGTCTTATGATTTTACATAATCCCTTTAGGGGGAGTGTCTTTTGTGT 46866
Qy 457 ----- 457
Db 46865 TTAAAGCAGTTTGTAGTACAGGAGTCTTGAACTTGGAAATTTTATTTCTCTTACGCTC 46806

```

QY 457 ----- 457
Db 46805 CCACTCAGCTGGATTACAGTCTTTGTGCCAACAGCCTAAGTGACATAGTTTTTAATTTT 46746
QY 457 ----- 457
Db 46745 TAGGACTGTTATTTTGGATCTTCAAACTCCATAAGTAGGAATGCAGATAGCACTTGTTA 46686
QY 457 ----- 457
Db 46685 GGTATTACAGACCTTAGTGCCACGACGACCCCTTCACCTTTGTTTTCAGCTTGTCTGACT 46626
QY 457 ----- 457
Db 46625 GCCTTCCGGAAGCGGCTCATGCTTTTAUCTTGGCTGGTTCAAGCTTTTATCTTGGCTGG 46566
QY 457 ----- 457
Db 46565 TTCAATGCTCAGCTGGTTATAGACAGATATAGAGCCTTCCCAATTACCAGGTGTTAAAAA 46506
QY 457 ----- 457
Db 46505 GAATGCTAAGGGTCTTTTGATGTGTGGCTAGGCCATAAAAAAGATACATGATTTCTACC 46446
QY 457 ----- 457
Db 46445 TAGCTTGGTTTGAAGGGTCTTTCAGGTTCTAAATAAAAGGTGTATATGTCTTTTAATTGGT 46386
QY 458 -----Glnal 459
Db 46385 TGTGAAGCAAAAGTGTCACTTCAAACCTTAATAATTTGGATTCTCAATATCCAGGT 46326
QY 459 aglylleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVa 479
Db 46325 GGGGATTAAATCTCTTCTGTGCACAGAGACTAGCGTCAGAGAAAGGGAACCCAGT 46266
QY 479 lLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysCysGI 499
Db 46265 GAAGAAA---GTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGCGCCTGTGA 46209
QY 499 uSerSerThrProSerTIPAlaSerAspHisAsnTyrAsnAlaValLysProGluLysTh 519
Db 46208 GAGCGACACCACTCTGGGCAAGTGCACCACTACATGCTGTGAGCCAGAGAGCC 46149
QY 519 rAlaAlaPro-----SerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyVa 536
Db 46148 AGAGAAAGCCCACTGCACTCTCGCCCACTTGTGAGTAAATGTACGTATCACCCAAAGGC 46089
QY 536 lGlyLeuLeuAspProSerArgSerPheTIPleAlaIleProTIPAlaCysProgly 555
Db 46088 TGGCTTTCCAGGCCCTCCCATC--ATCTGGGTGGCT--GCCTGGGGCTGTCTAGGA 46036

```

Search completed: May 1, 2004, 20:13:40
Job time : 5512 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 17:35:11 ; Search time 592 Seconds
(without alignments)
4294.234 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGDPSNEAPKAKPTS.....RSGFWIAIPWACPGIGVAAALC 562

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09787016/tunat_30042004_083703_5023/app_query.fasta_1.711
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNIT3=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09787016 @CNG 1 1 333 @runat 30042004_083703_5023
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
------------	-------	--------------	----	-------------

RESULT 1
US-09-925-301-327
; Sequence 327, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1398)

ALIGNMENTS

	1	1706	57.1	1764	9	US-09-925-301-327	Sequence 327, App
C	2	1092	36.5	645	15	US-10-029-386-25091	Sequence 25091, A
C	3	807	27.0	478	10	US-09-918-985-19127	Sequence 19127, A
	4	255	8.5	536	15	US-10-029-386-11381	Sequence 11381, A
	5	232	7.8	5975	10	US-09-873-367C-112	Sequence 112, App
	6	232	7.8	6256	15	US-10-037-270-369	Sequence 269, App
	7	232	7.8	6256	16	US-10-117-722-369	Sequence 269, App
	8	232	7.8	6948	10	US-09-930-213-384	Sequence 315, App
	9	214.5	7.2	2320	15	US-10-175-523-142	Sequence 142, App
	10	203	6.8	315	9	US-09-764-864-471	Sequence 471, App
	11	182	6.1	8346	13	US-10-263-929-20	Sequence 20, App
	12	181	6.1	1404	15	US-10-427-590-31	Sequence 31, App
	13	177	5.9	3787	10	US-09-930-213-384	Sequence 284, App
	14	176.5	5.9	7869	9	US-09-954-456-1921	Sequence 1921, App
C	15	176	5.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
	16	174	5.8	3609	9	US-09-917-800A-1715	Sequence 1715, Ap
	17	171.5	5.7	2502	13	US-10-342-887-1080	Sequence 1080, Ap
	18	171.5	5.7	2502	13	US-10-172-118-1080	Sequence 1080, Ap
	19	170.5	5.7	3183	13	US-10-263-929-50	Sequence 50, Appl
	20	170.5	5.7	4056	16	US-10-144-194A-85	Sequence 85, Appl
	21	170	5.7	3664	16	US-10-144-194A-73	Sequence 73, Appl
	22	169.5	5.7	5225	15	US-10-252-157-109	Sequence 109, App
	23	169	5.7	2031	10	US-09-823-187-23	Sequence 23, Appl
	24	169	5.7	2031	10	US-09-863-776-13	Sequence 13, Appl
	25	169	5.7	2031	15	US-10-203-708-5	Sequence 5, Appl
	26	169	5.7	2154	15	US-10-203-708-6	Sequence 6, Appl
	27	169	5.7	2390	15	US-10-102-524-1739	Sequence 1739, Ap
	28	168	5.6	2382	15	US-10-084-817-312	Sequence 312, App
	29	168	5.6	3174	13	US-10-282-122A-30251	Sequence 30251, A
	30	167	5.6	2004	10	US-09-863-776-15	Sequence 15, Appl
	31	167	5.6	2004	10	US-09-863-776-17	Sequence 17, Appl
	32	167	5.6	4530	13	US-10-302-172-881	Sequence 881, App
	33	166.5	5.6	3727	15	US-10-171-581-49	Sequence 49, Appl
	34	166	5.6	1938	15	US-10-106-698-1570	Sequence 1570, Ap
	35	166	5.6	2272	10	US-09-946-374-307	Sequence 307, App
	36	166	5.6	2272	12	US-10-015-395A-307	Sequence 307, App
	37	166	5.6	2272	13	US-10-208-915-345	Sequence 345, App
	38	166	5.6	2272	13	US-10-191-670-345	Sequence 345, App
	39	166	5.6	2272	13	US-10-201-858-345	Sequence 345, App
	40	166	5.6	2272	13	US-10-205-890-345	Sequence 345, App
	41	166	5.6	2272	13	US-10-208-024-345	Sequence 345, App
	42	166	5.6	2272	13	US-10-201-853-345	Sequence 345, App
	43	166	5.6	2272	13	US-10-174-581-345	Sequence 345, App
	44	166	5.6	2272	13	US-10-176-483-345	Sequence 345, App
	45	166	5.6	2272	13	US-10-176-749-345	Sequence 345, App

OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1758)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1759)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1762)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-925-301-327

Alignment Scores:
 Pred. No.: 5,618-148 Length: 1764
 Score: 1706.00 Matches: 317
 Percent Similarity: 99.37% Conservative: 0
 Best Local Similarity: 99.37% Mismatches: 1
 Query Match: 2 Indels: 2
 DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x US-09-925-301-327 (1-1764)

Qy 245 AsplleLysAspGluProGlyAsp-LeuGlyArgProLysProGluCysGluGlyT 264
 Db 2 GACATCAAGATGAGAGCTTGAGACTTTGGGCGGACCGAAGCTGGAATGGAGGTTA 61
 Qy 264 rAspProhnaLalaLeuTyrcyleCysArgGlnProhnaLalaLeuTyrcyleCys 284
 Db 62 CGACCCCAACGCCCTGATTCGATTTGGCGGAGCTTCAACACACGCTTTATGATTG 121
 Qy 284 sCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArg 304
 Db 122 CTGTGACCGCTGTGAGATGTTTCATGGCGATTTGTGGCATTTCTGAGGCTCGAGG 181
 Qy 304 YArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProhnaLalaLeuTyrcyleCys 324
 Db 182 GAGGCTTTTGGAAAGAAATGGGAAAGACTATATCTGCCCAACTGCAAGTCTCTCAAGT 241
 Qy 324 GlnAspGluTrpHisSerGluThraAlaAspGlnGluAlaLysTrpArgProGlyAs 344
 Db 242 GCAGGATGAGACTCATTCAGAAACGGGAGATCAGCAGAGAGCTTAATGGAGACCTGGAGA 301
 Qy 344 pAlaAspGlyThrAspCysThrSerileGlyThrleGluGlnLysSerSerGluAspG 364
 Db 302 TGCTGATGCGACCGATTTGACAGTATAGCAATAGCAGCAGAGTCTAGCGAAGACCA 361
 Qy 364 nGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysLys 384
 Db 362 AGGGATAAGGTTAGATTTGAGAAAGCTGCAATCCAGTGGCAAGAGAACTCAGAT 421
 Qy 384 ePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHi 404
 Db 422 CTTCAGGCTGTGATAGAGCGCTTGCTCCTCAAAATGTAATTGGCCCGGGTGTGTCA 481
 Qy 404 sValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLysHisAlaAla 424
 Db 482 CGTGGCGCA-CCGACTCGGTGTACTGCGAGTATGACTGTATCTCTCAACACAGCGCGCAGC 540
 Qy 424 aThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetL 444
 Db 541 GACATGAAGTTCTAAGCTCAGGTAAAGACAGAGCCCAAGCCCTAAAGAAAGATGAA 600
 Qy 444 sMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSe 464
 Db 601 GATGAGCCGAGAGAGCCAGCTTTCCGAAATGCGGTGCTCAGGAGGTTATTAATCTC 660
 Qy 464 rSerValHisLysArgProAlaProGluLysLysGluThrValLysLysAlaValVa 484
 Db 661 TTCTGTGCACAGAGACCGAGCTCCAGAAAAAAGAGACCAAGTGAAGAGGCGAGTGT 720
 Qy 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThProSe 504
 Db 721 GGTCTCCGCGAGTGAAGCACTCGGAAAGAAAGCAGCTTCTGAGAGCAGCAGCCCTC 780

Qy 504 rTPAlaSerAspHisAsnTyrcysAsnAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 781 GTGGCGGAGCGATCACAATTAATGATGATGATGATGATGATGATGATGATGATGATG 840
 Qy 524 oSerLeuLeuTyrcysCysMetTyrcysHisGlyValGlyLeuLeuAspProSerArgSe 544
 Db 841 GTCACTGTTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 Qy 544 rPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAlaLeuCys 562
 Db 901 TTTCTGATAGCATCCCTGGGCTGTCCAGGACTGGAGTTGCGAGTTTGTGT 955

RESULT 2

US-10-029-386-25091/c

; Sequence 25091, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AROMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: ArnoMax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 25091
 ; LENGTH: 645
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO ALL17379.14
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
 ; OTHER INFORMATION: SWISSPROT HIT: P23583, EVALUATE 2.50e-01
 ; OTHER INFORMATION: EST_HUMAN HIT: AL580669.1, EVALUATE 0.00e+00
 ; OTHER INFORMATION: NT HIT: G116193858, EVALUATE 0.00e+00
 US-10-029-386-25091

Alignment Scores:

Pred. No.: 9e-92 Length: 645
 Score: 1092.00 Matches: 214
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.53% Indels: 0
 DB: 15 Gaps: 0

US-09-787-016A-3 (1-562) x US-10-029-386-25091 (1-645)

Qy 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
 Db 643 ATGACCGACAAAGCGACCGGAGCAATGAGAGGACCTTAAGGCCATCAAAACCCACGAGC 584
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 Db 583 AAAGATTTCAGAAAAAATGGGGTTTTCGAGGACCACTATCGCCAGAGAGAGGGGCA 524
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 523 GGGGACGCGAGGCTGACCCACTGAGCGCGCCACCCACACAGCAGCAGCTGCGGCTGTGCC 464
 Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 463 CTGGCGCGCGTGGGAGGAGCCCAAGCGCACTGAGCGCGTGGAGGAGCTTCTGACCATTT 404
 Qy 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 403 CGCGCGCGCGCGGAGGAGCATGCTGTCTCTCCCTGGAGGATTTCTGGTGAAGCCACG 344
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 343 TCCTGCCCCCGCCACAGACGCGCGGAGAGCCTTCCGAGGCGCAGCGTGTAAAGCCTTCTGAG 284

Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 283 ACCAGAGCGGCCCCCTCTGCTTCACAGCTGTGAAGGACGACGAGCTCTTCGAA 224
Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 223 AAGGTGAAGGAGGGGATGACACACGATGACACCTCCGATAGTACAGCGATGGCTGACC 164
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGlnGlnGluProThrGluArgPro 180
Db 163 TTGAAGAGGCTTCAGAAATCGCTTCGAGGAGCGGGAACAGGAGCCCACTGAGAGGCC 104
Qy 181 LeuLysGlyLeuGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200
Db 103 CTGAAGGATCCAGATGCTGCGGAGAGCGCGGGAGGAGGTCCCGCGGAGCT 44
Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro 214
Db 43 GTGGGCTCCGAGGCCAGTGACACTGTGGAGGCGCTCTCTGCC 2

RESULT 3

US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19127
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(478)
; OTHER INFORMATION: n = A,T,C or G

Alignment Scores:
Pred. No.: 1,35e-65 Length: 478
Score: 807.00 Matches: 144
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.00% Indels: 0
DB: 10 Gaps: 0

US-09-787-016A-3 (1-562) x US-09-918-995-19127 (1-478)

Qy 215 SerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArg 234
Db 46 AGTAAGCAGGAGCGGAGAGCATCGGGGGTGTGTCCTCCAGGCTGGGAAGATGACAGA 105
Qy 235 GluSerLysLeuGluGlyLysAlaGlnAspLysAspGluGluProGlyAspLeu 254
Db 106 CAGAGTAAGTTGGAGGAAAGCGGCTCAGACATCAAGATCAGGAGCTCGAGACTTG 165
Qy 255 GlyArgProLysProGluCysGluGlyTrpAspProAsnAlaLeuTrpCysLeuArg 274
Db 166 GGCGAGCGAAGCGCTGAATGTGAGGGTTACACCCCAAGCCCTGTATGTCTATTCGCCG 225
Qy 275 GlnProHisAsnAspArgPheMetLeuCysCysAspArgCysGluGluTrpPheHisGly 294
Db 226 CAGCTCACACACACAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGGC 285
Qy 295 AspCysValGlyLysSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr 314

Db 286 GATTGTGTGGCATTCTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAGACTAT 345
Qy 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
Db 346 ATCTGCCCAACATGCACTTCTGCAAGTGCAGATGAGACTCATTCAGAAACGGCAGAT 405
Qy 335 GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
Db 406 CAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCAGATTGTACAGTATAGGA 465
Qy 355 ThrIleGluGln 358
Db 466 ACAATAGAGCAG 477

RESULT 4

US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US20030134704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Shazron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11381
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL17379.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUE 3.80e+00
; OTHER INFORMATION: NT HIT: G14769835, EVALUE 1.00e-78
; OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUE 2.00e-78

US-10-029-386-11381

Alignment Scores:
Pred. No.: 1.54e-14 Length: 536
Score: 255.00 Matches: 48
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.53% Indels: 0
DB: 15 Gaps: 0

US-09-787-016A-3 (1-562) x US-10-029-386-11381 (1-536)

Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 146 ATGAGCAGCAAAAGCGGACCCGAGCATGAGGAGCCTTAAGGCCATCAAAACCCACGAGC 87
Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 86 AAAGAGTTTCAGAAACATGCGGTTTTCGAAGGACCACTATGCGCAAGCGAGAGGGGCA 27
Qy 41 GlyAspAlaGluAlaAspProLeu 48
Db 26 GGGGACGCGGAGGCTGACCCACTG 3

RESULT 5

US-873-367C-112
; Sequence 112, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena

US-10-117-722-269

[illegible]

Db 1930 CCGACCAATGAAACCAACCAATAGTCACGTGAAGGAGAGCTTGAACACCCAGCCGCTTGAGCAT 1989
Qy 149 AspAspThrSerAspSerAspGlyLeuThrLeuLeuGlu-----Leu 164
Db 1990 -----TTTAAGGAAGAGGATAAATGAACCTGAAACAACTGAGAAGAACCTA 2037
Qy 165 GlnAsnArgLeuArgArgGlyArgGluGlnGluProThrGluArgProLeuLeuGlyLeu 184
Db 2038 CAACCCCGCCAAAGAAAGACGACCAAAAGTTTCTTTAGATGAGCCACCACTTGTTTCATT 2097
Qy 185 GlnSerArgLeuArgGlyArgGluGluGluProAlaGluThrValGlySerGlu 204
Db 2098 CCGGATAACATAGTACCATAGAGAGAA-----GGCTCT----- 2133
Qy 205 AlaSerAspThrValGluGluValLeuProSerLysGlnGluProGluAsnAspGlnGly 224
Db 2133 ----- 2133
Qy 225 ValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGln 244
Db 2134 -----GATCATAGTCTCTCATTTGAAAGCAAAATAT----- 2163
Qy 245 AspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyr 264
Db 2164 -----ATGTGCACTCCAGCAAGCAGTGT-----GGGTTT 2193
Qy 265 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 284
Db 2194 -----TGCAAAACCAACCATGCGCAACAGGTTTATGGTTGGC 2229
Qy 285 CysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAlaArgGly 304
Db 2230 TGTGGGAGATGTCATGACTGTTTTCATGTTGTTGTTGGTTAAAGTCTTTCTCAAGCA 2289
Qy 305 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 324
Db 2290 CAGCAGATGGCGGAGGAAGCAAAAGATATGCTGTGTAATAATGTGT-----GCT 2340
Qy 325 GlnAspGluThrHisSerGluThrAlaAsp-----GlnGlnGluAla 338
Db 2341 GAGAGAGCAAAAGACTGAAATAGTACTAGATCCAGATACITTTGGAACCAACAGTACAGTT 2400
Qy 339 LysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGln 358
Db 2401 GAATTCATAGTGGAGATAAACA-----ATGGAGTGTGAAGAGCTTGATTTATCAAAACAC 2457
Qy 359 LysSerSerGlu-----AspGlnGlyIleLysGlyArgIleGluLysAla 373
Db 2458 ACACCAATGATAGAACCAATATATAGATGATGATGAGTGAAGCAAGGTCAAAATTTTA 2517
Qy 374 AlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGly 393
Db 2518 AAACGGAGTCTGTGAAGGCAGA----- 2541
Qy 394 AlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyrCys 413
Db 2542 -----AATTCATCAGACTGT 2556
Qy 414 SerAsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLys 433
Db 2557 AGAGATAATGAAATTAATAAATGCGAGTACTCTCTCTGTAAGATGGGCAACCAAGTT 2616
Qy 434 GluGlnLysProLysProLysGluLysMetLysMetLysProLysProSerLeuPro 453
Db 2617 TTACTCTGGAGATCTCTCAGAGAAAGAAAGTGAATAATACCGAAAGAGTCTCAACTGTT 2676
Qy 454 LysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGlu 473
Db 2677 ACTTGC-----ACAGAGAGAAAGCTTCAAAACCAAGTACTCATGAGAGCAAGAG 2727
Qy 474 LysLysGluThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGly 493

Db 2728 ATGAAAAGAGAGAAAGTTGAAAAGGAGTG-----CTTAAT 2763
Qy 494 LysGluAlaAlaCysGluSerSerThrProSer 504
Db 2764 GTACATCTGCTGCTCTGCTTCCAAAGCCTTCT 2796
RESULT 9
US-10-175-523-142
; Sequence 142, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHPTS)
; FILE REFERENCE: 3235/J1795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-175-523-142
Alignment Scores:
Pred. No.: 4,9e-10 Length: 2320
Score: 214.50 Matches: 95
Percent Similarity: 38.38% Conservative: 42
Best Local Similarity: 26.61% Mismatches: 98
Query Match: 7.18% Indels: 122
DB: 15 Gaps: 20
US-09-787-016a-3 (1-562) x US-10-175-523-142 (1-2320)
Qy 224 GlyValValSerGln---AlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAla 242
Db 24 GGGTCTGTGAGGAGGTGCGCGGAGCGAGATATGAGGGA----- 65
Qy 243 AlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGlu 262
Db 66 -----GATGGTTCAGACCCAGAGCTTCAGATGCCGGGAGGAGCAGC---AAGTCCGAG 116
Qy 263 GlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMet 282
Db 117 AATGGGAGAAATGCGCCCATCTACTGTCATCTCCGCAAAACCGGACATCACTGCTTCATG 176
Qy 283 IleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAla 302
Db 177 ATCGGGTGTGACAACTGCAATGAGTGTTCATGGGACTGCTCGGAGTCACTGAGAAG 236
Qy 303 ArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeu 322
Db 237 ATGCCCAAGCCCATC-----CGGAGTGTACTGTCCGGAGTGC----- 275
Qy 323 GlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArg---- 341


```

RESULT 10
US-09-764-864-471
; Sequence 471, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P7223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (894)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

RESULT 11

```

US-10-263-929-20
; Sequence 20, Application US/10263929
; Publication NO. US20040067535A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Jaeseob
; APPLICANT: Galant, Ron
; TITLE OF INVENTION: Alzheimer's Disease Linked Genes
; FILE REFERENCE: LSD-07417
; CURRENT APPLICATION NUMBER: US/10/263,929
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 8346
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-263-929-20

Alignment Scores:
Pred. No.: 2,28e-06 Length: 8346
Score: 182.00 Matches: 54
Percent Similarity: 33.73% Conservative: 80
Best Local Similarity: 18.93% Mismatches: 99
Query Match: 6.09% Indels: 125
DB: 13 Gaps: 9

US-09-787-016A-3 (1-562) x US-10-263-929-20 (1-8346)
QY 90 ProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGluThr 109
Db 6990 CCAAGTTACCTATCCAAATTCAGCAAAGCAGTGTCTGTCAGACTCCACAGATTCAGAAATG 7049
QY 110 AlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 129
Db 7050 GGTTCACAGTCGACGGCAGCAGTGTGCAAGAGCAGTGTGCAAGGGTTTCACCAACTCAGGCA 7109
QY 130 ThrAlaValLysGluArgProAlaSer----- 138
Db 7110 TCAGCAGCAAAACAGAAACAGCAACAGATAGAGAAATTAAGCCGTGAACACACCCCTCCAAGC 7169
QY 139 -----SerGluLysValLysGlyGlyAspAsp-HisAspAspThrSerAs 153
Db 7170 TTCTAATCAAGTGAATCAATTCAGAAACAGGTGTGTGATGAAGCATATGCTGTATAGA 7229
QY 153 pSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgArgGluArgG1 173
Db 7230 A-----CATTTAAACAGAAAAAGAG 7250
QY 173 uGlnGluProThrGluArgPro-----LeuLys 182
Db 7251 CATGACTCCAGCTGAAAGAGAGAGAGAAATCAAGATGATGTCTGTAAACAGGTGATGA 7310
QY 182 sGlyIle-----GlnSerArgLeuArgLysLysLysArgGluG1 195
Db 7311 GTATATTTTGATAGATAGATAAGAGAAAGAAACAGGCAGCAGCAAAACAGGAGCGTGA 7370
QY 195 uGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuProSe 215
Db 7371 AGAGAGTGTGGAGCAGAAACAGTAGCAAGCAGAGATGCCACTAGCTGTGCTCTT 7430
QY 215 rLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgG1 235
Db 7431 CAGGCACAAAGACGACTCGAGCCGAGATCTCTGAAGAAGAGAGAGCACTCTGTGACAGGA 7490
QY 235 userLysLeuGlu-----GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAs 253
Db 7491 TCTGCAAAATGAAGTGCAGCAGAGAGCTGAAGAGAGACCTGGAATAATTAAGAAAGAAAAGA 7550
QY 253 pLeuGlyArg----- 256
Db 7551 CCTGATGCAGTGTGGCTCAGGCCACAGACAGTAGCTGCACCTGCCCTCCAGTGCACCACT 7610
QY 257 ----ProLysProGluCysGluGlyTyrAspPro----- 266

```

Db 194 ACCACTCGCAGCGCGCGCGCGCGCTGCTGGACACCGTGAACCTACCCCATCCACATGA 253
Qy 87 ArgSerMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAsp 106
Db 254 AGAACCTGTCCC---TCAAGGAGCTGCAGACGCTCTCCGACGAGCTGGCTCCGACGTCA 310
Qy 107 AlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGln 126
Db 311 TCTTCCACGTCTCCAAAGACCGCGC-----GCCACTCGGGTCCAGCC 352
Qy 127 SerAlaSerThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyLys 146
Db 353 TCGGGTGTGTGAGCTCACGTGCGGTGCTACTAGTCTTCAACACCCCGC----- 403
Qy 147 AspHisAspAspThrSerAspSerAspSerGlyLeuThrLeuLysGluLeuGlnAsn 166
Db 404 -----AGGACCAAGCTCTCTGGGAGCTGGCCACCACTGCTTACCGC-CACAAG 450
Qy 167 ArgLeuArgArgLysArgGluGluGluProThrGluArgProLeuLysGlyLys 186
Db 451 ATTCTGACGGCGCGCGCGATAGATGCCGAGATGCGGACAGACCAACGCGCTTCCGGC 510
Qy 187 ArgLeuArgLysLysArgGluGluGluProAlaGluThrValGlySerGluLys 206
Db 511 TTGTGTC-----NAGCGCTCCGAGAGCGAGTACGACAGCTTCGGCAGCGCGCACAGC 561
Qy 207 AspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyVal 226
Db 562 TCCACCACCATCTCCGCGC-CCTCGGATGCGCGTGGAGGAGCACTCAAGGGCGCGAA 620
Qy 227 SerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAsp 246
Db 621 GAAACAACGTGTGGCGGTGATTGGGACGGCGCATGACGCGCGCGCAGGC-----GTA 674
Qy 247 LysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyrAspPro 266
Db 675 CGAGCGATGAACAACGCGCGCTACTCGA-----CTCGGACATGAT--- 716
Qy 267 AsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAsp 286
Db 717 -----CGTGATCTCAACGACCAACAGCAGGCTGCTGCGCGAC 755
Qy 287 ArgCysGluGluTrpPheHisGlyAspCysValGlyLysSerGluAlaArgGlyArgLeu 306
Db 756 -----GGCAGCGCTCGACGCGCGCGCGCGCGCGCGCGCGCT 794
Qy 307 LeuGluArgAsnGlyLysAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAsp 326
Db 795 CAGCGCGC----- 803
Qy 327 GluThrHisSerGluThrAlaAspGlnGlnAlaLysTrpArgProGlyAspAlaAsp 346
Db 804 -----CCTCAGCAAGCTCAGTCCAGCGCGCGCTCAGGAGCTGAGGAGGTGGCAA 857
Qy 347 GlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyLe 366
Db 858 GGGAGTGAAGCAAGCAAAAT-----CGGCGGCTCGGTGCACGAGATCGCGCGCAA 905
Qy 367 LysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheGln 386
Db 906 GGTGACGAGTAGAGCGCC-GCGCATGATCAGCG----- 937
Qy 387 ProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValala 406
Db 938 -----GCTCCGGGTGCTGCTCTTCGAGGAGCTCGGCGCTGATT---ACATCG 982
Qy 407 GlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAlaAlaThrMet 426
Db 983 GCCCGCTGACGCGCC-----ACAACTGAAGACC 1012
Qy 427 LysPheLeuSerSerGlyLysGluGlnLysProLysPro----- 439
Db 1013 TCATCACCATCTCTGGGAGGTCAAGGGCACCAGACCCCGCGCGCGTGTCTCATCCATG 1072

Qy 440 -----LysGluLysMetLysMetLysProGluLysProSerLeuProLysCys 455
Db 1073 TCATCACCAGAAAGCGCGGTACCCCTACCCCGGAGCGCTCCGACAAAGTACCAAC 1132
Qy 456 GlyAlaGlnAlaGlyLysIleSerSerValHisLysArgProAlaProGluLysLys 475
Db 1133 GGTGGCAAGTTCGATCCGCGCACCGGAGGAGTTCAGGGTCCGGCCCAAGACGCGCTT 1192
Qy 476 GluThrThrValLysLysAlaValValProAlaArgSerGluAla-----Leu 492
Db 1193 CCTACACA-----ACTACTTCGCGAGCGCTCATAGCCGCGG 1234
Qy 493 GlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerAspHisAsnTyrAsn 512
Db 1235 GGCAGACACAGATGCTGCGCATTCACCGGCCATGGGGCGGAGCGGGCTCAACT 1294
Qy 513 AlaValLysProGluLysThrAlaAlaProSerProSerLeuLeuTyrLysCysMetTyr 532
Db 1295 ACT-----TCTCCGCGCTTCCCAACC----- 1318
Qy 533 HisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAla 552
Db 1319 -----GGGCTTCCAGTCCGATTCGCT-AAACAGAACCGCGTAAACCTTCCGCGCGC 1371
Qy 553 CysProGlyLeuGlyValAlaAlaLeuCys 562
Db 1372 CTGCGCGCAGGGGTAAACCTTCTGTC 1401

RESULT 13
US-09-930-213-284
; Sequence 284, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHE-NITSE, OLEG
; APPLICANT: GRIPS, MARTIN
; APPLICANT: HELLMER, MARTIN
; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SERS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBR-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 284
; LENGTH: 3787
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-930-213-284

Alignment Scores:
Pred. No.: 2,56e-06 Length: 3787
Score: 177.00 Matches: 108
Percent Similarity: 36.18% Conservative: 49
Best Local Similarity: 24.88% Mismatches: 162
Query Match: 5.92% Indels: 119
DB: 10 Gaps: 20

US-09-787-016A-3 (1-562) x US-09-930-213-284 (1-3787)
Qy 189 ArgLysLysArgArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThr 208
Db 141 CGGCTCCCGCGCGCGCCAGCAGCGCGCGCTGATGTGTGACGCGCGCGCGCGCGA 200
Qy 209 ValGluGly---ValLeuProSerLysGlnGluProLysAsnAspGlnGlyValValSer 227

```

QY      527  -----LeuTyrIyaCyMetTyrHisIeuG 535
Db      1223 ACAGCGAGATGTTCTTTGTCGACCGAGTGCACAAATGCTACAAAGTCATCGTCACGACGAG 1282
QY      535 lyValGlyLeuLeuAappProserArgSerPheTrpIle 547
Db      1283 GCCAGACCTCTTCATCCCTCAGGC-----TGGATC 1314

RESULT 14
US-09-954-456-1921
; Sequence 1921, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1921
; LENGTH: 7869
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1921

```

[illegible]

QY 90 rovalserleuGluaspSerGlygluProThrSerCysProAlaThrAspAlaGluThrA 110
DB 4046 CCCTACCCGAGGAGGAGGAGGAGGAGATG-----CAAGCAGGAGGAGGAGGAGG 4099
QY 110 laSerGluGlySerValGluSerAlaSerGluThrArg-----SerGlyProGlns 127
DB 4100 CTGAGAGGGTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4159
QY 127 erAlaSerThrAla-----VallysGluA 135
DB 4160 ACAGCAGCAATAGTCTCTGAGACCGAAACCAAGGAGGAGGAGGAGGAGGAGGAGGAG 4219
QY 135 rgProAlaSerSerGluLysVallys-----GlyGlyAspAspHisAspA 150
DB 4280 CAGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4338
QY 150 spThrSerSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgA 170
DB 4339 -----GCTGATGATGAGGATGATGCT-----CACCTGGAGTCCCAAGAGAAA 4391
QY 170 rGlySerGluGlnGluProThrGluArgProLeuLys-----GlyleGln----- 185
DB 4382 AGGAGCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4441
QY 186 -----SerArgLeuArgLysLysArgGluGluGluGluProAlaGluThrValG 202
DB 4442 CTTTTTTAGATGCTAATATGAGAGAGTAGG----- 4473
QY 202 lySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGlu-----ProGluA 221
DB 4474 --GAAAGATAAGGATAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4531
QY 221 snAspGlnGlyValValSerGln-----AlaGlyLysAspAsp----- 233
DB 4532 ATGACACGTCGCTGT 4591
QY 234 -----ArgGluSerLysLeuGluGlyLys----- 241
DB 4592 CCCACACTAAGGAGAGTAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 4651
QY 242 -----AlaAlaGlnAspIleLysAspGluGluProGlyAspLeuG 255
DB 4652 TGGATCTGGAACCTGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4711
QY 255 lyArgProLysProGluCysGluGlyTyArgProAsnAlaLeuTyrcysileCysArg- 274
DB 4712 AGGGGCGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4756
QY 275 -----GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysG 289
DB 4757 CCGTCGAGAGTTACACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4816
QY 289 luGluTrpPheHisGlyAspCysVal----- 297
DB 4817 ATGCGTCAAGACATAATAGCCCTATCTCTCGTTTCAGTCTCACCCAGGAGGAGGAGGAGGAG 4876
QY 298 --GlyleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrcileCysp 317
DB 4877 GTTCGCTCAGCAGTCCCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4933
QY 317 roAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnG 337
DB 4934 CAGAAACAGGAGTCCCTGTGTCGAGCCCTCTATGTCAGAAATGAGGAGGAGGAGGAGGAGGAG 4993
QY 337 luAlaLysTrpArgProGlyAspAlaAsp-----GlyThrAspCysT 351
DB 4994 ATGTG-----CCTTCCGTTATCAGACCACTCTCAGCAGGAGGAGGAGGAGGAGGAGGAG 5044

QY 351 hrSerileGlyThrile-----GluGlnLysSerSerGluAspG 364
DB 5045 GTACCTTGGGACGATGTGAGAGCACCCTGAAACCTATGTAGAACCCAGGAGTACGACT 5104
QY 364 lnglyleLysGlyArgileGluLysAlaAlaAsnProSerGlyLysLysLysLysLysLys 383
DB 5105 CCACGATGGCGGCGAGCATC-----TGTGGGAACAGCTCTTCCAGAGAGGAGGAGGAGGAG 5158
QY 384 -----IlePheGlnProValI 389
DB 5159 ACGGTGGGCTGTCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5218
QY 389 leGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProA 409
DB 5219 TGGCCAGCATGGCGGAGCAGC-----TGACGATCATGTCAGCAGAGA 5257
QY 409 spSerValTyrcysSerAsnAspCysIleLeuLysHis-----AlaAlaAlaThrMetLysP 428
DB 5258 GCAGCGTCCAGCTGTCTGCACTGAGCATCAAGTCACTCAGAGTGGTGGTGAGA 5317
QY 428 heLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProG 448
DB 5318 GGCTCTCCCATTAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5377
QY 448 luLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisL 468
DB 5378 CGCCGCCA----- 5389
QY 468 ysArgProAlaProGluLysLysGluThrThrValLysLysAlaValValProAlaA 488
DB 5390 CACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5424
QY 488 rgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSerTrpAlaSerA 508
DB 5425 CAGCGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5474
QY 508 sphIsanTyrcysAlaValLysProGluLysThrAlaAlaProSerProSerLeuLeuT 528
DB 5475 -----CCTTCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5511
QY 528 yrLysCysMetTyrcysHisLeuGly----- 535
DB 5512 ATGAATAACAGTTCCTATGATCATGAGATACAGAGATCTGGAAGCAGCT 5571
QY 536 -----ValGlyLeuLeuAspProSerArgSerPheTrp 546
DB 5572 GGGAAACATAAGTATCTATGAGAGGATTCAGGAGGATTTTGG 5612

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 0.0343 Length: 9025608
Score: 176.00 Matches: 140
Percent Similarity: 34.47% Conservative: 72
Best Local Similarity: 22.76% Mismatches: 218
Query Match: 5.89% Indels: 189
DB: 15 Gaps: 24

US-09-787-016A-3 (1-562) x US-10-156-761-1 (1-9025608)

QY 53 ProGluGlnLeuLeuSerLeuArgSerGlyArgGlnProLysArgThrGlu 72
DB 5319462 CTTGGCAGTATCAGGCGCCGAGGAGCCGTCAGAGCTGATCAGCTCGCTGGC 5319403

QY 73 Arg-----ValGluGlnPheLeuThrIleAlaArgArgGlyArgSerMetPro 90
DB 5319402 CGGCGTGGCGCCGACACATCCGTCATCGTGGCGAGTCTCTCGACA-TGATGTCAG 5319344

QY 91 ValSerLeuGluAspSerGlyGluProThr----- 100
DB 5319343 GTGACACCTTCGCGTGCAGCTCTCGACACCCCGCTCGTGGTGGGCGGTCGCGGAGA 5319284

QY 100 ----- 100
DB 5319283 ACCAGAACCTGTCAGACACAGCGGCCATGACAGCGCCGCGCGCGCTCGCACC 5319224

QY 101 -----SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSer 117
DB 5319223 TGGGGCGCCAGCTGCTCCGCGGACGAGGCGGTGGCGGAGTTCGCGCCGCGCTCGC 5319164

QY 118 AlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAla 137
DB 5319163 GCGACACCCGCTAGTCTCCGCGGCGAGCTTCGCGAGTCTACTCGACGAGACTTCG 5319104

QY 138 SerSerGluLysValLysGlyLysAspPheAspAspThrSerAsp----- 153
DB 5319103 TCAAGCGCGCACCGGCTCAAGTGGCTGAAGAGATCTTCTCATCATCTGCTGGCCCTCG 5319044

QY 154 -----Ser-AspSerAspGlyLeuThrLeuLys 162
DB 5319043 CGGTATCGGCGCGGAGACGTACGTGGCTTACCGCTGGAGCGCAGACGAGTA----- 5318992

QY 162 sGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgProLeuLys 182
DB 5318991 ----CTAGTTCGCGCGCAGCGGAGCACGTCCGCTGTACCGGGCATCAGCAGGACC 5318936

QY 182 sGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThrValGlu 202
DB 5318935 TGGCCTGGGTCTCCTCTCGAA-----GG 5318912

QY 202 ySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAs 222
DB 5318911 TGCAGAGGACACCCGAGATCGAATCACTCACTGCTGCGCGGTACCGGATCAGCAGACGAG 5318852

QY 222 pGlnGly-ValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysA 242
DB 5318851 TCAAGCGCAGCATCGCGGAGGCGGTCTCAAGACCGCGAGTCAAGAGTCAAGCACTCG 5318792

QY 242 laalaglnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysG 262
DB 5318791 CGGTACAGGC-GTCCGCTGTAGAA-----GGACGCTCTGCGCG----- 5318752

QY 262 luGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheM 282
DB 5318751 -----TGCTTC-----CGAGAGCCAGCAACACGCCAA----- 5318725

QY 282 etIleCysCysAspArg-----CysGluGluTrpPheHisGlyAspCysValGlyIleS 300

Search completed: May 1, 2004, 21:43:26
Job time : 5588 secs

5318724 -----GACCGCGAGCGCGAGCGAGCTGGGACCAACGCGGACCAACAGACCTCT 5318676
QY 300 exCluaalaargGlyArgLeuLeuArgAsnGlyGluAspTyrIleCysProAsnCysT 320
DB 5318675 CTGACGTCCAGGCGACGCTCGTGCC-----GACCCCGAGCCCGTCGAGCTCGGACCG 5318622

QY 320 hrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGln----- 336
DB 5318621 TGTCTCTCAAGTCCAAAGACCGACCTCCACACCGCGCCCGCCAGCTCTTCGAGAA 5318562

QY 337 -----GluAlaLysTrpArgProGlyAspAlaAspG 347
DB 5318561 GACGAGAACCTGCTCTGCTGCGGTAGCAGTAGAAGACCGCTGAGGG----- 5318512

QY 347 lyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleL 367
DB 5318511 -----CCCTGTACACCATGAGCAGTAGTACGACACGCGCGACGACACGCTCGACCA 5318457

QY 367 yGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLys 387
DB 5318456 TGGCTCGATCGCGCGCACGAGCGCGCACACCGAGTTGGCTGCTGGTGTTCGCG 5318397

QY 387 roValIle-GluAla-----ProGlyAlaSerLysCysIle 398
DB 5318396 TGTCTATCCCGGTCTTCGGGTAGCCACAGCTGCGCTGCGCTGCTCGACGAAATGC----- 5318340

QY 399 GlyProGlyCys-----CysHisValAlaGlnProAspSer 410
DB 5318339 CGCGCGGCTGCTGAGCTACGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318280

QY 411 ValTyrCysSerAsnAspCysIleLeuLysHis-AlaAlaAlaThrMetLysPheLeuSe 430
DB 5318279 GCGTACGCAAGTTCAGCGCGACCGTACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 5318234

QY 430 rSerGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLys 449
DB 5318233 -GGCGACGCTGCTCAAGCGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318175

QY 450 ----Pro-SerLeuProLysCysGlyAlaGluAlaGlyLysLysLysLysLysLysL 468
DB 5318174 TGTCTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318119

QY 468 ys-----ArgProA 471
DB 5318118 GCGTGGGATCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318059

QY 471 laProGluLysLysGluThrThrValLysLysAlaValValValProAlaArgSerGluA 491
DB 5318058 CAGCGCTACACCTACATCTCCATGTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318008

QY 491 laLeuGlyLysGluAlaAlaCysGluSerSerThr-----P 503
DB 5318007 CCGCTGGTTC-CGGGCTCGGCGCCACACATCTACGCGCGCAGATCTGATCTCGATTCC 5317949

QY 503 roSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAlaAlaProS 523
DB 5317948 CGGCTGGGACCTCC-----AGCCGCGAGTTCGGGAGATCATCTCGCGGT 5317898

QY 523 erProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAspProSerA 543
DB 5317897 CTCTTCG-----CGGCTA 5317883

QY 543 rGserPheTrpIleAlaIleProTrpAlaCysPro 554
DB 5317882 CTTCTGTCAGCGCGAGCGCTGCGCTCGCCCTCGCCA 5317848

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model
Run on: May 1, 2004, 17:18:53 ; Search time 124 seconds
(without alignments)
2515.181 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPDSREAPKAIKPTS.....RSFWIAIPWACPGVGAALC 562

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09787016/runat 30042004 083703 4995/app_query.fasta_1.711
-DB=Issued Patents NA -CFMT=fastap -SUPPLY=p2n.rni -MINMATCH=0.1 -LCOPCL=0
-LCOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @cgn 1.1.85 @runat 30042004 083703 4995 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	232	7.8	6256	4	US-09-620-312D-269
2	181	6.1	1404	4	US-09-857-556A-31
3	174.5	5.8	8931	3	US-09-051-019-1
4	170.5	5.7	4056	4	US-10-164-595-55
5	167	5.6	3228	4	US-09-252-991A-1816
6	167	5.6	3546	4	US-09-252-991A-2243
7	163.5	5.5	2433	4	US-09-489-039A-2577
8	161.5	5.4	4248	4	US-10-164-595-53
9	161.5	5.4	6605	1	US-08-769-309A-4
10	161.5	5.4	6605	4	US-08-994-570-4
11	161.5	5.4	6608	4	US-09-220-132-58
12	161.5	5.4	6755	3	US-08-931-999-4

13	161	5.4	1032	4	US-09-257-179-21	Sequence 21, Appl
14	160	5.4	2344	3	US-08-893-852A-2	Sequence 2, Appl
15	158.5	5.3	12842	3	US-09-105-537-30	Sequence 30, Appl
16	158.5	5.3	36778	3	US-09-105-537-5	Sequence 5, Appl
17	158.5	5.3	38506	3	US-09-320-878-19	Sequence 19, Appl
18	158.5	5.3	38506	4	US-09-141-908-1	Sequence 1, Appl
19	158.5	5.3	38506	4	US-09-657-440-19	Sequence 19, Appl
20	157	5.3	3240	4	US-09-489-039A-331	Sequence 331, App
21	157	5.3	5173	4	US-08-801-308-2	Sequence 2, Appl
22	154	5.2	1887	4	US-09-252-991A-15843	Sequence 15843, A
23	154	5.2	2103	4	US-09-252-991A-15876	Sequence 15876, A
24	154	5.2	5200	4	US-08-978-277A-3	Sequence 3, Appl
25	152.5	5.1	3900	4	US-09-252-991A-14550	Sequence 14550, A
26	152.5	5.1	4403765	3	US-09-103-840A-2	Sequence 1, Appl
27	152.5	5.1	4411529	3	US-09-103-840A-1	Sequence 1, Appl
28	152	5.1	4048	4	US-09-976-594-1053	Sequence 1053, Ap
29	150.5	5.0	11236	1	US-07-853-913-1	Sequence 1, Appl
30	150	5.0	9439	3	US-09-058-489-89	Sequence 89, Appl
31	150	5.0	9551	1	US-08-056-200-93	Sequence 93, Appl
32	150	5.0	9551	2	US-08-800-644-93	Sequence 93, Appl
33	149.5	5.0	2050	4	US-09-620-312D-761	Sequence 761, App
34	149.5	5.0	8146	4	US-09-976-594-725	Sequence 725, App
35	149.5	5.0	11917	4	US-09-566-921-32	Sequence 32, Appl
36	149	5.0	8438	1	US-07-945-283-1	Sequence 1, Appl
37	149	5.0	12508	4	US-09-655-270A-1	Sequence 1, Appl
38	149	5.0	12523	4	US-09-651-941-1	Sequence 1, Appl
39	149	5.0	12523	4	US-09-955-597-1	Sequence 1, Appl
40	148	5.0	2302	4	US-09-620-312D-915	Sequence 915, App
41	147	4.9	2262	4	US-09-252-991A-1060	Sequence 1060, Ap
42	147	4.9	4257	2	US-08-690-473-1	Sequence 1, Appl
43	147	4.9	4257	4	US-09-259-821A-1	Sequence 1, Appl
44	147	4.9	4257	3	US-08-843-659-1	Sequence 1, Appl
45	147	4.9	5121	4	US-09-252-991A-15189	Sequence 15189, A

ALIGNMENTS

RESULT 1
US-09-620-312D-269
; Sequence 269, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghaast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT FILING DATE: 2000-07-19
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-01-21
; PRIOR FILING DATE: 2000-01-21
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA


```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257)..(5443)
US-09-620-312D-269

Alignment Scores:
Pred. No.: 4,96e-10 Length: 6256
Score: 232.00 Matches: 94
Percent Similarity: 36.43% Conservative: 63
Best Local Similarity: 21.81% Mismatches: 154
Query Match: 7.76% Indels: 120
DB: 4 Gaps: 17

US-09-787-016A-3 (1-562) x US-09-620-312D-269 (1-6256)
Qy 92 SerLeuGluAsp----SerGluGluProThrSerCysProAlaThrAspAlaGluThrAla 110
Db 1106 TCTTTGAGTGAATGATCAACGCTCATCCTGCTGCTTGAAGAACCTCATCATCTCTGCA 1165
Qy 111 SerGluGlySerValGluSerAlaSerGluThrArg----SerGlyProGlnSerAlaSer 129
Db 1166 CAACCTGGACATGATCATCTAGCCAGAAACAGTCTCATAGCCTCAGCACAGGCC 1225
Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLys----GlyGlyAspHis 148
Db 1226 CCAGCAATGAAACCAATAGTCACTGAGGAGAGAGCTTGAACACCCAGGCGTTGAGCAT 1285
Qy 149 AspAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGlu-----Leu 164
Db 1286 -----TTTAAAGAGAGAGTAACTGAACCTGAAATAAATAAAGAGAGAGCTTA 1333
Qy 165 GlnAsnArgLeuArgGlyArgGluGlnGluProThrGluArgProLeuLysGlyIle 184
Db 1334 CAACCCGCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1393
Qy 185 GlnSerArgLeuArgGlyLysArgArgGluGluGluGluProAlaGluThrValGlySerGlu 204
Db 1394 CCAGATAACATGATACCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429
Qy 205 AlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGly 224
Db 1429 ----- 1429
Qy 225 ValValSerGlnAlaGlyLysAspArgGluSerLysLeuGluGlyLysAlaAlaGln 244
Db 1430 -----GATCATAGCTCTCTCATTTGAAAGCAATAT----- 1459
Qy 245 AspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyr 264
Db 1460 -----ATGTGACTCCAGCAGCAGCTGT---GGGTTT 1489
Qy 265 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 284
Db 1490 -----TGCATAAACCACATGTCCTGTGTAATGTTGT-----GCT 1525
Qy 285 CysAspArgCysGluGluThrPheIleGlyAspCysValGlyIleSerGluAlaArgGly 304
Db 1526 TGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1585
Qy 305 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 324
Db 1586 CAGCAGATGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1636
Qy 325 GlnAspGluThrHisSerGluThrAlaAsp-----GlnGlnGluAla 338
Db 1637 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1696
Qy 339 LysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGln 358
Db 1697 GAATTCATAGTGGAGATGATAAACA---ATGGAGTGTGAAGAGCTGATGATTAACAAAC 1753
Qy 359 LysSerSerGlu-----AspGlnGlyIleLysGlyArgIleGluLysAla 373
```

RESULT 2

```

US-09-857-556A-31
; Sequence 31, Application US/09857556A
; Patent No. 6558915
; GENERAL INFORMATION:
; APPLICANT: Rebecca E. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng
; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
; FILE REFERENCE: BB1290
; CURRENT APPLICATION NUMBER: US/09/857,556A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-857-556A-31
```

```

Alignment Scores:
Pred. No.: 1,57e-06 Length: 1404
Score: 181.00 Matches: 127
Percent Similarity: 33.68% Conservative: 65
Best Local Similarity: 22.28% Mismatches: 244
Query Match: 6.06% Indels: 138
DB: 4 Gaps: 20
```

US-09-787-016A-3 (1-562) x US-09-857-556A-31 (1-1404)

```

Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 14 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 73
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
Db 74 GCGTGTCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133
```

```

QY 61 LeuArgArgSerGlyArgGlnPro-----LysArgThrGluArgValGluGlnPhe 77
Db 134 AGACGCCGAGGAGGAGTGTCCGCGATTCTCGCTGCTCGGAGGAGGAGCAGAGT 193
QY 78 LeuThrIleAlaArgArgArgGlyArg----- 86
Db 194 ACCATCTCGAGCGCGCGCGCGCGCTCTCTGGACACCGTGAACCTACCCCATCCATGA 253
QY 87 ArgSerMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAsp 106
Db 254 AGAACCTGTGCC---TCAAGGAGCTGCAGCAGCTCTCCGACGAGCTCGCTCCGAGTCA 310
QY 107 AlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGln 126
Db 311 TCTTCCACGCTCTCCAGACCGCGG-----GCCACCTCGGCTCCAGCC 352
QY 127 SerAlaSerThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyGlyAsp 146
Db 353 TCGGCGTCTCGAGCTCACCGTCCGCGTCACTACGTCTTCAACACCCCGC----- 403
QY 147 AspHisAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGluGlnAsn 166
Db 404 -----AGGACAAAGCTCTCTGGGACGTGCGGCCACCACTGTATCCCG-CACAAAG 450
QY 167 ArgLeuArgArgLysArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSer 186
Db 451 ATTCTGACGGCGCGCGCGCATTAAGATGCGGACGATCGCGCAGACCAACGCGCTGTCCGCG 510
QY 187 ArgLeuArgLysArgGluGlnGluProAlaGluThrValGlySerGluAlaSer 206
Db 511 TTCCGT-----AAGCGCTCCGAGACGAGTACGACAGCTTCGCGCAGCGGCCACAGC 561
QY 207 AspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyValVal 226
Db 562 TCCACCAACCATCTCCCGCGC-CCTCGGGATGCGCGTCCGGGAGGACCTCAAGGGCGCGAA 620
QY 227 SerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyValAlaGlnAspIle 246
Db 621 GAACAAAGTGTGGCGGTGATTTGGGACGCGGCGCATGACGCGCGGCGCAGC-----GTA 674
QY 247 LysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyrrAspPro 266
Db 675 CAGGGCGATGACACACCGCGGTACTCTCA-----CTCGGACATGAT--- 716
QY 267 AsnAlaLeuTyrrCysIleCysArgGlnProHisAsnAspArgPheMetIleCysCysAsp 286
Db 717 -----CGTGATCTCTCAACGACAAACAGCAGGTGTCTGCTGCGGAC 755
QY 287 ArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeu 306
Db 756 -----GCGGACGCTCGACGCGCGCGCGCGCGCGCGCGCGTGTGCGCGGT 794
QY 307 LeuGluArgAsnGlyGluAspTyrrIleCysProAsnCysThrIleLeuGlnValGlnAsp 326
Db 795 CAGCGCGCGC----- 803
QY 327 GluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArgProGlyAspAlaAsp 346
Db 804 -----CCTCAGCAAGCTCAGTCCAGCGCGCGCTCAGGAGAGCTGAGGAGTGCGCCAA 857
QY 347 GlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIle 366
Db 858 GGGAGTGACGAACCAAT-----CGCGGCGTGTGTCACGAGATCGCGGCCAA 905
QY 367 LysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLeuLysIlePheGln 386
Db 906 GTTGACGAGTACGCGCC-CGCGCATGATCAGCG----- 937
QY 387 ProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAla 406
Db 938 -----GCTCCGAGTGTGCTCTTCCAGGAGCTCGGCTGTATT---ACATCG 982
QY 407 GlnProAspSerValTyrrCysSerAsnAspCysIleLeuLysHisAlaAlaThrMet 426

```

```

Db 983 GCGCGCTCGCGCGC-----ACACATTGACGACC 1012
QY 427 LysPheLeuSerSerGlyLysGluGlnLysProLysPro----- 439
Db 1013 TCATCACCATTCTTCGGGAGGTCAAGGCGCACCAAGACACCGCGCGGTGTCTATCCATG 1072
QY 440 -----LysGluLysMetLysMetLysProGluLysProSerLeuProLysCys 455
Db 1073 TCATCACCAGAAAGCGCGCGCTACCTTACCGCGAGGAGCTTCCGACAGTACCAC 1132
QY 456 GlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLys 475
Db 1133 GGTGGCAAGTTCGATCCGGGCGCGGAGGAGTTCAGGTTCAGGTTCGCGCAGAGCGCTT 1192
QY 476 GluThrThrValLysLysAlaValValProAlaArgSerGluAla-----Leu 492
Db 1193 CCTACACA-----ACTACTTCGCGGAGCGCTCTATAGCCCGAGGCGG 1234
QY 493 GlyLysGluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnTyrrAsn 512
Db 1235 GGCAGACAGCAAGATCGTGGCATTCACCGCGCATTCGGGCGCGGCGGCGGCTCAACT 1294
QY 513 AlaValLysProGluLysThrAlaAlaProSerProSerLeuLeuTyrrLysCysMetTyrr 532
Db 1295 ACT-----TCCTCCGCGCTTCCCAACC----- 1318
QY 533 HisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAla 552
Db 1319 -----GGGCTTCCAGTGTGATCCGCT-AAACGAAAGCGGTAAACCTTCCGCGCGC 1371
QY 553 CysProGlyLeuGlyValAlaAlaLeuCys 562
Db 1372 CTGGCGCGCAAGGGGTAAACCCCTTCTGCG 1401

```

RESULT 3

```

US-09-051-019-1
; Sequence 1, Application US/09051019
; Patent No. 6103229
; GENERAL INFORMATION:
; APPLICANT: KAHMANN, Regine and QUADBECK-SEGER, Claudia
; TITLE OF INVENTION: Regulatory gene from Ustilago maydis
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kell & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, Pentium processor
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Wordperfect version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,019
; FILING DATE: 31-MAR-1998
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8931 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-051-019-1

```

```

Alignment Scores:
Pred. No.: 9,46e-05 Length: 8931
Score: 174.50 Matches: 129
Percent Similarity: 36.62% Conservative: 94
Best Local Similarity: 21.18% Mismatches: 204
Query Match: 5.84% Indels: 186

```


LOCATION: (189)...(2528)

OTHER INFORMATION:

US-10-164-595-55

Alignment Scores:

Pred. No.: 6,48e-05 Length: 4056
Score: 170.50 Matches: 132
Percent Similarity: 35.68% Conservative: 76
Best Local Similarity: 22.64% Mismatches: 235
Query Match: 5.70% Indels: 141
DB: 4 Gaps: 24

US-09-787-016A-3 (1-562) x US-10-164-595-55 (1-4056)

QY 7 ProSerAsnGluGluAlaProLysAlaIleLeuProThrSerLysGluPheArgLysThr 26
Db 861 CCCTCAGAGATCTGATTCCTACCCAGAGAG-----AAGAAGAGAGAGAGAGAG 914
QY 27 TrpGlyPheArgThrThrIleAlaLysArgGluGlyAlaGlyAspAlaGluAlaAsp 46
Db 915 GACATTTTCGAGATTTCAGTGCC-----941
QY 47 ProLeuGluProProProGlnGlnGlnLeuGlyLeuSerLeuArgSerGlyArg 66
Db 942 CCACTGATCCCTTATCCACTCATCAAGGA-GGATATAATGCTATAGAAATGGAAGA 1000
QY 67 GlnProLysArgThrGlu-----ArgValGluGlnPheLeuThrIleAlaArgArg 84
Db 1001 AGACAAAGAGAGCCTGATATCTCGAGATGATCAACAAATTCAGAGACACATAGAAGACT 1060
QY 85 GlyArgArgSerMetProValSerLeuGluAspSerGly-----GluProThrSerCys 102
Db 1061 GGAAGAGAGAGAGGCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 103 ProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArg 122
Db 1118 -----AGAAAG 1171
QY 123 SerGlyProGlnSerAla-----SerThr 130
Db 1172 AAGGAG 1231
QY 131 Ala-ValLysGluArgProAlaSerSerGluLysValLysGlyGlyAspAspHisAspAs 150
Db 1232 GCAG 1291
QY 150 pThrSerAspSerAspSerAspLysLeuThrLeuLysGluLeuGlnAsnArgLeuArgAr 170
Db 1292 CCAG 1351
QY 170 GlysArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLys 190
Db 1352 GAAA-----GAAGCTGTTATCAGAGAGGC---CTTAAGAATTGGGAATCAGAGAGAGAA 1405
QY 190 slys-----ArgArgGluGluGlyProAlaGluThrValG 202
Db 1406 GAAACCCGGGATATGAG 1465
QY 202 ySerGluAla-----SerAspThrValGluGlyValLeuProSerLysGlnGluPr 219
Db 1466 CAAGAGAGCTAAACGACTAAAGAGATCTTAGAGAGACTATGATGATGATGATGATGATG 1525
QY 219 oGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGl 239
Db 1526 CAAATATTACAG 1585
QY 239 uGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPr 259
Db 1586 AGCA-----GATGACGA-----GATGACGA-----GATGACGA-----GATGACGA 1618
QY 259 oGluCysGluGlyTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAs 279
Db 1619 CGAGCTTGAG-----1628

QY 279 nArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyTl 299
Db 1628 -----1628
QY 299 eSerGluAlaArgGlyArgLeuLeuGluArgGlyGluAspTyIleCysProAsnCy 319
Db 1629 -----GAATCAGGAGAGCGCTTCTGCGAGAGGCGATCCAGAT-----CCAGATCC 1675
QY 319 sThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
Db 1676 AGAGCTCCAGAGGATGATGACCAAGAG-----GCTCAGAGGCGCAGGCGAGCAATAAA 1729
QY 339 sTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
Db 1730 GCAAGAGCGCAATCAGAGAGAGAGAGAG-----GAAGAGAA 1765
QY 359 sSerSerGlnGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
Db 1766 GCAAG 1825
QY 379 sLysLysLysLys-----IlePheGlnProValIleGluAlaProGlyAlaSerLysCys 398
Db 1826 GCCTTGTCTGAAACCTACTCTGAGGCGCCATCAGCTCTGCTCCATCTGTCTCTGCTGCCAG 1885
QY 398 eGlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCys-- 417
Db 1886 TGCGAAT-----GCAACACCTAACACTCTGGGATGAGTCTCCCTGTGG 1930
QY 418 -----IleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLys 436
Db 1931 TATTATTATCTCATGAA-----AACTCACAGATCAACAGCA 1969
QY 436 sPro-----LysProLysGluLysMetLysMetLys-----ProGluLysPr 450
Db 1970 ACCTGAG 2029
QY 450 oSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisArgPr 470
Db 2030 TGCTCAGCTTAATCTCTGAG 2089
QY 470 oAlaProGluLysLysGluThrThrValLysLysValValValValProAlaArgSerGl 490
Db 2090 TGAGATGAG 2149
QY 490 uAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHisAs 510
Db 2150 TGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2209
QY 510 nTyrAsnAlaValLysProGluLysThrAlaAlaPro-SerProSerLeuLeuTyLysC 530
Db 2210 TAAGAGTCTCTGAG 2262
QY 530 yMetTyHisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaLys 550
Db 2263 -----CCCTGGATGCTGCTATTG-----TGGATCTATACT 2293
QY 550 rofTrp 551
Db 2294 GATGG 2298

RESULT 5

US-09-252-991A-1816

Sequence 1816, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 1816
;; LENGTH: 3228
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1816

Alignment Scores:
Pred. No.: 9,33e-05 Length: 3228
Score: 167.00 Matches: 140
Percent Similarity: 36.93% Conservativeness: 86
Best Local Similarity: 22.88% Mismatches: 217
Query Match: 5.59% Indels: 172
DB: 4 Gaps: 30

US-09-787-016A-3 (1-562) x US-09-252-991A-1816 (1-3228)

QY 4 LysGlyAspProSerAsnGluAlaPro---LysAlaIleLysProThrSerLysGlu 22
DB 1553 AAATGCCACCGCTCGAGCAGGAAGCCACCGCTGAGTCCGACCCGACCCCTGGTCC 1612
QY 23 PheArgLysThrTyrGly-PheArgArgThrThrIleAlaLys----- 36
DB 1613 GCCAGAGAGCAGCAGTCAAGACCGTCTCCAGCAGCCGCGCACACACACCGAAG 1672
QY 37 -----ArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProProProGlu 54
DB 1673 CACCGGTGAGCGCGCAGCGCGATGCCGAGCCGAGCGCTGT-----CCAGGCGCTGG 1726
QY 54 nGlnGlnLeuGlyLeuSerLeuArgArgSerGly-----ArgGln-ProLysArg 71
DB 1727 TGAAGTCCCTGCTGGCTGCTGTTGAGGCAAGATCACTGCGCCAGCTGCTGAAA 1786
QY 71 hrGluArg-----ValGluGlnPheLeuThrIleAlaArgArgGlyArgArg 88
DB 1787 CCAGCAAGCGCGTCCGAGCGCCAAACCGCGCAGCAGAGGTGCGCAACGCGCGCCAGC 1846
QY 88 erMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAla 108
DB 1847 AG-----AACCGCGCGCGATGCGCGATGCGCGATGCGCGCGCGCGCGCGCGCG 1896
QY 108 luThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyPro----- 125
DB 1897 AAACCGCG---CGAGGAGC-GTGACAGAGGTCAACCGCGCAGAGAGCGCGCGCGCGCG 1952
QY 126 -----GlnSerAlaSerThrAlaValLysGluArgProAlaSerS 139
DB 1953 GAACCGCGAAGCGCGAGCG 2011
QY 139 erGluLysValLysGlyLysAspAspHisAspThrSerAspSerAspGlyL 159
DB 2012 AGGAGCGC----- 2019
QY 159 euThrLeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGlu 179
DB 2020 -----CAGCGCGCGAGAGCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 2062
QY 179 rg-----ProLeuLysGlyIleGlnSerArg---LeuArgLysLysArgArg 194
DB 2063 GCGAAGCG 2122
QY 194 luGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuP 214
DB 2123 GCGAAGAGCG 2182
QY 214 roSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAsp 234
DB 2183 AACCGGTGAGACCG 2230
QY 234 rgGluSerLys-----LeuGluGlyLysAla-----A 243

RESULT 6
US-09-252-991A-2143/c

DB 2231 GCCAGGCTCGAGCGCGCGCGCGCTGAGAGCCGAGGAGCATTCGCCAGACGACGAGCGCTGG 2290
QY 243 laGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluG 263
DB 2291 AGCAGGACGAGCAGGAGGATACCGATGCGGAG-----CGCCCGCGCGCGCGCTCCGCG 2344
QY 263 lyTyrrAspProAsnAlaLeuTy-CysIleCysArgGlnProHisAsnAsnArgPheMetI 283
DB 2345 GCCAG-----CGTGTGCGCAGCAGCAGCGCGCG----- 2370
QY 283 leCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAla 303
DB 2371 -----GAACGCCAGCGCGAG-----GTACAGCGCGAGGTGTGAAGGC---AGCAGCGCA 2416
QY 303 rgGlyArgLeuLeuGluArgAsnGlyCyluAspTyrrIleCysProAsnCysThrIle--- 321
DB 2417 CC-----GATATGCGCGCGCGCGCTGAAACACCGCTGCGCAG 2452
QY 322 -----LeuGlnValGlnAspGlu----- 327
DB 2453 CCGCGCGCGTCCGGTATCGCTGTCAGAGGAGCGGTAGAACAAACGTGAGCAAG 2512
QY 328 -----ThrHisSerGluThrAlaAspGlnGlnGluAlaLysThrPargProGlyAsp 345
DB 2513 CCGCGCGCCTTACCGAGCGAGCTGCCAGCGAAACACCGCA-----A 2554
QY 345 laAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGln 365
DB 2555 GCGATGAGACCGACCGCTGCGACCGAGCGCGTGTGAAACCCAGCAGCGCGACGAGCG 2614
QY 365 lyIleLysGlyArg-----IleGluLysAlaAlaAsp 376
DB 2615 CCAATACCGCGGAAACCGCGACATCGAAGCGCGCGTGTGCGTGTGCGGAGCG 2674
QY 376 roSerGlyLysLysLysLysIlePheGlnProValIleGluAlaProGlyAlaSerL 396
DB 2675 AAGCGACGACGACCGCTGCTGTCGCGCAAGCACCCTGAGAGAGCTCCCTTCGCCAGCG 2734
QY 396 ysCysIle----- 398
DB 2735 AAAGCGTGAAGCGCGCGAGACCGCGAGCGCGCTGCAACCGCGCAACGCGCGCGCG 2794
QY 399 -----GlyProGlyCysCysHisValAlaGlnProAspSerValTyrrCysSerA 415
DB 2795 AAGAGTGTGCGCTCCGCGTCCGCGTCCGAGTAGCAGCCCTTAGCGAGCGCGCGACCG 2854
QY 415 snAspCysIleLeuLysHisAlaAla-----AlaThrMetLysPheLeuSerS 431
DB 2855 AGGAGCGACCG 2914
QY 431 erGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProS 451
DB 2915 ACCCAGCGGAAAGCGCTCGCTGCAACGCGAGCGCGCGCTGTGCGACGCGAGCGCGCG 2974
QY 451 erLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProA 471
DB 2975 CAGCAGCGAGCGCGCGCTCAGCGCGCT-----CCCGCGTCCGAGGAGATCCCGG 3025
QY 471 laProGluLysGluThrThrValLysLysAlaValValProAlaArgSerGluA 491
DB 3026 CTGTAGCGCGAGGAGCG 3076
QY 491 laLeuGlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerHisAsnT 511
DB 3077 CT-----GAGAAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3124
QY 511 yrAsnAlaValLysProGluLysThrAlaAla 521
DB 3125 AGGAAGCGGTACAGCG 3156

```

; Sequence 2143, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2143
; LENGTH: 3546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2143

Alignment Scores:
Pred. No.: 0.000107 Length: 3546
Score: 167.00 Matches: 140
Percent Similarity: 36.93% Conservative: 86
Best Local Similarity: 22.88% Mismatches: 217
Query Match: 5.59% Indels: 172
DB: 4 Gaps: 30

US-09-787-016A-3 (1-562) x US-09-252-991A-2143 (1-3546)
Qy 4 LysGlyAspProSerAsnGluGluAlaPro---LysAlaIleLysProThrSerLysGlu 22
Db 1814 AAATGGCCACCGTGCAGCAGCAAGAACCCAGCCGCGTGCAGTCGACCCGCAACCCCTGGTCC 1755
Qy 23 PheArgIleThrTrpGly-PheArgArgThrThrIleAlaLys----- 36
Db 1754 GCCAGGAGCAGCAGTCAAGACCTGCTCCCGAGCGCCGACCGCAACACACCGAAG 1695
Qy 37 -----ArgGluGlyAlaGlyAspAlaGluAlaAspProLysGluProProGlu 54
Db 1694 CACCGTGCAGCGCCAGCCGATGCGCGAGCGCGCGTCTCCCGAGCGCCGACCGCAACCGCCGCG 1641
Qy 54 nGlnGlnLeuGlyLeuSerLeuArgArgSerGly-----ArgGln-ProLysArgT 71
Db 1640 TGAAGTCCTGCTGGCTGTTCGAGCGCAGGATCACTGCGCGCAAGCTCTCTGCTGAA 1581
Qy 71 hrGluArg-----ValGluGlnPheLeuThrIleAlaArgArgArgArg 88
Db 1580 CCAGCAAGCGCGTCCGAGCGCCAAACCCCGAGCAGCGTCCGCAACCGCGCCGCG 1521
Qy 88 erMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaG 108
Db 1520 AG-----AACCGCGCGCGATGCCCGCGATGG-CAATCGCCCGCAGCAAGAGCGC 1471
Qy 108 luThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyPro----- 125
Db 1470 AAAACCGCG---CGAGGAGC-GTGCAGAGCGTCAACCGCGCGAGAGCGCGCGCAACGCC 1415
Qy 126 -----GlnSerAlaSerThrAlaValLysGluArgProAlaSerS 139
Db 1414 GAACCGCGAAGAGCGCAGCAACCGCGCGCAAGAGCGCGCC-GAGGTCGCGCTCGCG 1356
Qy 139 erGluLysValLysGlyAspAspHisAspThrSerAspSerAspGlyL 159
Db 1355 AGAGCGCG----- 1348
Qy 159 euThrLeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluA 179
Db 1347 -----CAGCGCGCGAAGCGCGCTGAGAGCGCGCGCGCGGACACCCC 1305
Qy 179 rg-----ProLeuLysGlyIleGlnSerArg---LeuArgLysLysArgArg 194
Db 1304 GCGAAGAGCGCAACCGCGCGAGCGCGCGAGGTCGCGAGGAGCAGCAGCGAGCGCGCGCC 1245

```

```

Qy 194 luGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuP 214
Db 1244 GCGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
Qy 214 roSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspA 234
Db 1184 AACGGGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
Qy 234 rgGluSerLys-----LeuGluGlyLysAla-----A 243
Db 1136 GCAGGCTCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1077
Qy 243 laGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluG 263
Db 1076 AGCAGGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1023
Qy 263 lyTyraAspProAsnAlaLeuTyrcysIleCysArgGlnProHisAsnAsnArgPheMetI 283
Db 1022 GCCAG-----CGTGTGCAGCAACCGCGCG----- 997
Qy 283 leCysCysAspArgCysGluGluTrpPheHisLysAspCysValGlyIleSerGluAlaA 303
Db 996 -----GACGCCAGCGCGAG---GTCAGCGCGCGAGGTGGAAGGC---AGCGAGCGCA 951
Qy 303 rgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCysThrIle----- 321
Db 950 CC-----GATAATGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 915
Qy 322 -----LeuGlnValGlnAspGlu----- 327
Db 914 CCGCGCGCGTCCGCGTATCGTGTCCGAGCGAGCGCGTAGAAGCAACAGTGGAGCAAG 855
Qy 328 -----ThrHisSerGluThrAlaAspGlnGlnAlaLysTrpArgProGlyAspA 345
Db 854 CCGCGCGCACTACCGCGAGGTGCGCAGCGAACCACCGCA-----A 813
Qy 345 laAspGlyThrAspCysThrSerIleGlyThrIleGlnLysSerSerGluAspGlnG 365
Db 812 GCGATGAGCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Qy 365 lyIleLysGlyArg-----IleGluLysAlaAlaAsp 376
Db 752 CCAATACCGCGCAACCGCGCATCGAAGCGCGGTGACCGTCAGCGTGTTCGGGAGCG 693
Qy 376 roSerGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerL 396
Db 692 AAGCGCAGCAGACCGCTGCTGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
Qy 396 ysCysIle----- 398
Db 632 AAAGCGTGAAGCGCGCAAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
Qy 399 -----GlyProGlyCysCysHisValAlaGlnProAspSerValTyrcysSerA 415
Db 572 AAGAAGTTCGCTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCG 513
Qy 415 snAspCysIleLeuLysHisAlaAla-----AlaThrMetLysPheLeuSerS 431
Db 512 AGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
Qy 431 erGlyLysGluGlnLysProLysProLysProLysMetLysMetLysProGluLysProS 451
Db 452 ACCCAGCGGAAACCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
Qy 451 erLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerValHisLysArgProA 471
Db 392 CAGCAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
Qy 471 laProGluLysLysGluThrThrValLysLysAlaValValProAlaArgSerGluA 491
Db 341 CTGTAGCGCGCGAGGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 291

```


Qy	520	-----AlaaLPrSerProSerLeuLeuTyrHisL	534
		:::	
Db	384	CAGCCTCCGGCGCTGCACCCACCCTCATGTC	337
		:::	
Qy	534	euglyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaalleProTrpAlaCysP	554
		:::	
Db	336	CTGTGTAGCGTCTC-----CGCCGTAAACACAGCAGATGTTCCAGCAGGTGCGC	289
		:::	
Qy	554	roGlyLeuGlyValAlaalaLeu	561
		:::	
Db	288	CAGCCCAGGGAAGCGCACGGTT	266

RESULT 8

```

US-10-164-595-53
; Sequence 53, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53:
; LENGTH: 4248
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (189)..(2720)
; OTHER INFORMATION:
; US-10-164-595-53

```

Alignment Scores:	
Pred. No.:	0.000428
Score:	161.50
Percent Similarity:	34.06%
Best Local Similarity:	21.54%
Query Match:	5.40%
DB:	4
	Length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	Gaps:
	4248
	139
	75
	246
	150
	24

US-09-787-016A-3 (1-562) x US-10-164-595-53 (1-4248)

Qy	2	AspAspIysGlyAspProSerAsnGluGluAlaProIlysAlaIleYsProThrSerLys	21
Db	999	GAAGACAAAGAGACCTGATCTCGAGAGATCAGCAAA	1037
Qy	22	GluPheargIysThrTTPdGlyPheArgArgThrThrIleAlaLysArgGluGlyAlaGly	41
Db	1038	---TTCAGAGACACA---CATTAAGAACTCGAAGAAGAAAGCGCAAAAG	1082
Qy	42	AspAlaGluAlaaspProLeuGluProProProGlnGlnLeuGlyLeuSerLeu	61
Db	1083	GAATAAGAAACAGCGAAATTGAGAAAGACGGAGAGAAAGAGAGGGAGCGTGAAGG	1142
Qy	62	ArgArgSerGlyArgGlnProLysArgThrGluArg-ValGluGlnPheLeuThrIleAl	81
Db	1143	GAACGAGAAAGCGGAGAACCGGAAACGA---GAAAGGGAAAGAGAAACGTGAACGAGAAAG	1199
Qy	81	aArgArgArgGly-----	85
Db	1200	GAGAAAGACGGGCGGGACCGAAGACGGATAGGACCGTGACCGACAAAGAGAGA	1259
Qy	86	-----ArgArgSerMetProValSerLeuGluAspSerGlyGluProTh	100
Db	1260	GACCGGATCGGGATCGAGAGAGATCGTCGCGGATAGAGAAAGAGCTCAGATCGT	1319
Qy	100	rSerCysProAlaThrAspAlaGlu-----ThrAlaSerGluGlySe	114
Db	1320	AATAAGATCGCGATCGTCAAGAGAAAAAAGACAGATCGTGAAGGGGACGAGACGG	1379
Qy	114	rValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValLysGl	134

[illegible]

Db 2234 TTCTGTGAAGAGAAAGAACTACTGTAGATAGTGTCTTTAACAAATTTGAGGATGAAGA 2293
 QY 474 slysgluThrVallyLysAlaValValValProAlaArgSerGluAlaLeuGlyLY 494
 Db 2294 CAGTGATGAGTACCCCGAAAGAAAGAACTGGTTCCTTGGATTTATGCTGAAGATGATAA 2353
 QY 494 sgluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaVa 514
 Db 2354 AAATGCAACCAAGGCACTGTAACACACTGAAGAAAGCGTAAACACATTAAGAGTCTCAT 2413
 QY 514 llyProGluLysThrAlaAlaPro-SerProSerLeuTyrLysCysMetTyrHisL 534
 Db 2414 TGAGAAATCCCTACAGCAACCTGAGCTCTTCCTCTATC-----C 2455
 QY 534 euGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrp 551
 Db 2456 CCTGGATTGCTCTATTG-----TGGATTCTATCTACTGATGG 2490

RESULT 9

US-08-769-309A-4
 ; Sequence 4, Application US/08769309A
 ; Patent No. 5741890
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klaucek, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/769,309A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5741890and, Greta E.
 ; REGISTRATION NUMBER: 35,302
 ; REFERENCE/DOCKET NUMBER: 27866/33451
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-474-6300
 ; TELEFAX: 312-474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6605 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 192...5531
 ; US-08-769-309A-4

Alignment Scores:
 Pred. No.: 0.000832 Length: 6605
 Score: 161.50 Matches: 133
 Percent Similarity: 30.51% Conservative: 76
 Best Local Similarity: 19.42% Mismatches: 221
 Query Match: 5.40% Indels: 255
 DB: 1 Gaps: 28

US-09-787-016A-3 (1-562) x US-08-769-309A-4 (1-6605)

QY 3 AsplyGlyAspProSerAsnGluAlaPro-----LysAlaIleLysProThrSer 20
 Db 987 GAGGAGGAGAGAGAAACAAAGAAACCTAGCAAGCTCTGCAGAACTCTCCGACTAGT 1046
 QY 21 -----LysGluPheArgLysThrTrp-----Gly 28
 Db 1047 CCGTGACCAAGTGAACAGGATCAACCTTCAAAAAATTTCTACTCAAGGTTGGGCCGCG 1106
 QY 29 PheArgArgThrThr----- 33
 Db 1107 TGGCGCAAAAGACCACTTTTCAGSAGCCGAGGAGGATGAAGTGAAGCTTTCAGAGAG 1166
 QY 34 -----IleAla 35
 Db 1167 AAAAAGGAACAGAGCCAGAAAAAGTAGACACAGAAAGACGGAAGGAGGAGGTTGCC 1226
 QY 36 LysArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProPro----- 51
 Db 1227 TCCGAGAACTGACCGCTCCGAGCAAGCCACCACACAGGAGCCGCGGAGAAAGTGCAC 1286
 QY 52 -----ProGlnGlnGlnLeuGly 58
 Db 1287 GAGCCCGCTTATCAGCTGAATATGAGAAAGTTGAGCTGCGCTCAGAGGAGCAAGTC--- 1343
 QY 59 LeuSerLeuArgArgSerGlyArgGln-----ProLysArg 70
 Db 1344 -----AGTGCTCCAGGACCTTCTCGAAGAGAACTGCTCCGCTCGGTTGGCG 1388
 QY 71 ThrGluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgArgSerMetPro 90
 Db 1389 ACAGAGTGTGTGATGAGAAATAGAAAGTCCACCAAGAGAGGTTGTGGCCGAGATCCAC 1448
 QY 91 ValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAla---GluThr 109
 Db 1449 GTCAGCACCCGTGGAG-----GAGAGAACCAAGACGAGAAACGAGGTCGAGAGAAC 1502
 QY 110 AlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 129
 Db 1503 GCAGGCTGTGCGCAGCTGAAGATTTGTTGGAATGATGATGAGAACTCAGAACTCAGAA 1562
 QY 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyLysAspHisAsp 149
 Db 1563 CTTGCC---AAGGAGCTGTGGAAGCTCAAGAAAGCTGTGTTCCGAGAGACCTTACA 1619
 QY 150 AspThrSerAspSerAspGlyLeuThrLeuLysGlu----- 163
 Db 1620 CAGGAGCTGACCTCAGTCTCTGATGAGAGGCTGTGCCAAACCCCGAAGGCGTTGTG 1679
 QY 164 -----LeuGlnAspArgLeuArgArgLysArgGluGlnGluPro----- 176
 Db 1680 AGTCAGGTGGAAATGCTGTCTATCACAGGAGAGATGAAGGTGCGAGGAGTCCACTAAG 1739
 QY 177 -----ThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLysArgArg 193
 Db 1740 AAGCTTTTACCAGCACTGGCTTAAAGAGCTTCTGGAAGAAACAGAAAGGGAAGAA 1799
 QY 194 GluGluGly-----ProAlaGluThrValGlySer 203
 Db 1800 GGAGGAGGAGACGAGGAATCAGGGAGACACACTCAGGTTCCAGCCGATTTCTCCGAG 1859
 QY 204 GluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu----- 220
 Db 1860 CAGGAGGAGCAAAAGGGCGAGAGCTCTGCTCATCTCCCTGAGGAGCCGAGGAGATCA 1919
 QY 221 -----AsnAspGlnGlyValValSerGlnAlaGlyLys 231
 Db 1920 TGTCTGAAAGAGGCTTAGCCGAGGTGAGCAGGATGGGAGGAGCTGAAGAGGAGTACT 1979
 QY 232 AspAspArgGluSerIleLeuGluGly----- 240
 Db 1980 TCCGATGAGAGAGAAAAAGAGAGAGGTTGCTACTCCCTGGGCGATCATTCATCAAAAG 2039

Db 1287 GAGCCCCGGTTATACGTGAATATAGAAAGTTGAGTGCCTCCAGGAGCAAGTC--- 1343
Qy 59 LeuSerLeuArgArgSerGlyArgin-----ProLysArg 70
Db 1344 -----AGTGGCTCGCAGGACCTTCGAGAGAGAAACCTGCTCGGTGGCG 1388
Qy 71 ThrGluArgValGluGlnPheLeuThrIleAlaArgAArgGlyArgSerMetPro 90
Db 1389 ACAGAAAGTGTGTGATGAGAAATAGAGTCCACCAAGAGAGGTTGTGCGCCAGATCCAC 1448
Qy 91 ValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAla---GluThr 109
Db 1448 GTACACACCTGGAG-----GAGAGAACCCAGAGAGCAGAAACGAGGTGGAAGAAACA 1502
Qy 110 AlaSerGluGlySerValGlnSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 129
Db 1503 GCAGGGTCTGTGCCAGCTCAAGAAATGTGTTGAATGATGATGAGACCTTCAGAGAACCCGAA 1562
Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyAspAspHisasp 149
Db 1563 CTGTC-----AAGAGCTGTGAGCTCAAGAAACGTGTGTTCCGAGAGACCTTACA 1619
Qy 150 AspThrSerAspSerAspGlyLeuThrLeuLysGlu----- 163
Db 1620 CAGGGAGCTGACCTCAGTCTGATGAGAAAGTGTCTCCAAACCCCGAAGCGGTGTG 1679
Qy 164 -----LeuGlnAsnArgLeuArgAArgGlyArgGluGlnGluPro----- 176
Db 1680 AGTGAGGTGGAAATGCTGTATCAGCAGAGAGATGAGGTGCAGGAGATCCACTAAG 1739
Qy 177 -----ThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysArg 193
Db 1740 AAGCTTTTACCAGCACTGGCTTAAAAAGCTTCTCGAAAGAAACAGAAAGGGAAGA 1799
Qy 194 GluGluGly-----ProAlaGluThrValGlySer 203
Db 1800 CGAGGAGGAGCAGGAAATCAGGGAGCACATCAGTTCCAGCCGATTCCTCCGACAGC 1859
Qy 204 GluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu----- 220
Db 1860 CAGGAGGACCAAAAGCGGAGAGCTGCTCCTCATCCCTGAGGAGCCCGAGAGATCAG 1919
Qy 221 -----AsnAspGlnGlyValValSerGlnAlaGlyLys 231
Db 1920 TGTCTGAAAAGGGCTTAGCCGAGTGCAGCAGGATGGGAGCTGAAAGAGGAGCTACT 1979
Qy 232 AspAsnArgGlnSerLysLeuGluGly----- 240
Db 1980 TCCGATGGHAGAGAAAPAAAGAGAGAGGTGTCACTCCTCGGCATCATTCAAAAGATGGTG 2039
Qy 240 ----- 240
Db 2040 ACGCCCAAGAGCGTGTAGACGGCTTCGAAAGTGATAAAGAGAGATGAGCTGGACAG 2099
Qy 241 -----LysAlaAlaGlnAspIleLysAspGluGlu 250
Db 2100 GTCAAGAGCGCTACCTGTCTCCACCGAGAGCAGCAGCTCTGAAATGCAAGAGAAATG 2159
Qy 251 ProGlyAspLeuGlyArgProLysProGluCysGluGlyTyraAspProAsnAlaLeuTy 270
Db 2160 AAAGGAGCGTGGAGAGCCAAAGCCGAA----- 2189
Qy 271 CysIleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGlu 290
Db 2190 -----GACCAAAAGCCCAAG-----GTGGATACCTCAGTATCT 2222
Qy 291 TrpPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsn 310
Db 2223 TGGAGAGCTTTAATTGTGTGGATCATCCAGAAAGAGCAGG-----AGAAGGTCC 2276
Qy 311 GlyGluAspTyrlleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSer 330
Db 2277 TCTTCTGAT-----GAGAGAGGGGAGCAAAAGCA 2306

Qy 331 GluThrAlaAspGlnGlnGluAlaLysTrpArgProGlyAspAlaAsp---GlyThrAsp 349
Db 2307 ATGGAGGAGACACCCAGAAAGCTGATGAGCCGCGAAAGACAAAGAGCGGGACAGAC 2366
Qy 350 CysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArg 369
Db 2367 GGGATCCTTCTGCTGCTCC-----CAAGAACATGATCATCCAGCGCAGGCA---AGTTCCTCC 2417
Qy 370 IleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProValIle 389
Db 2418 CCGAGCAAGCTGGAGCCCTTACC----- 2441
Qy 390 GluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAsp 409
Db 2442 GAGGGGAGCGCGTTC----- 2459
Qy 410 SerValTyrcysSerAsnAspCysIleLeuLysHisAlaAlaThrMetLysPheLeu 429
Db 2460 -----ACCTGGGAGTCAATTTAAAGGTTA 2483
Qy 430 SerSerGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLys 449
Db 2484 GTACG-----CCAGAAAAAATCAAACTCCAGCTCGAGAG 2522
Qy 450 ProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArg 469
Db 2523 AAAAGCGAAGACTCCATA---GCTGGGTCTGGTGTAGACATTCCTCCACAGACTCAA 2579
Qy 470 ProAlaProGlnLysLysGluThrThrValLysLysAlaValValProAlaArgSer 489
Db 2580 CCCGTTAAAGAGAAATCTGGTCTCAATCAAGAGTTTATCTCTGACGAGAGAGAA 2639
Qy 490 GluAlaLeuGlyLys-----GluAlaAlaCysGluSerSerThrProSerTrpAlaSer 507
Db 2640 AGCCAGATGGAAACAAGAACAGCCCTCTTGAAGACGCGGCGCAACAGCGGCCAAC 2699
Qy 508 AspHisAsn-----TyraAsnAlaValLysPro 516
Db 2700 GAGATGACTCTGATGTCCCGCGCTGCTCTCTGTGTGATGATGCTGTAGAAAGG 2759
Qy 517 GluLysThrAlaAla 521
Db 2760 GAGAAATGGAGGCA 2774

RESULT 11

US-09-220-132-58

; Sequence 58, Application US/09220132

; Patent No. 6506607

; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; FILE REFERENCE: 07334-074001

; CURRENT APPLICATION NUMBER: US/09/220,132

; PRIOR FILING DATE: 1998-12-23

; PRIOR APPLICATION NUMBER: US 60/079,303

; PRIOR FILING DATE: 1998-03-25

; PRIOR APPLICATION NUMBER: US 60/068,821

; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 58

; LENGTH: 6608

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-220-132-58

Alignment Scores:

Pred. No.: 0.000833

Score: 161.50

Percent Similarity: 30.51%

Best Local Similarity: 19.42%

Length:

Matches: 133

Conservative: 76

Mismatches: 221

32 ThrThrIleAlaLysArgGluGluValAlaGluValAspAlaGluAlaAspProLeuGluProPro 51
592 GCAACCAAGTCTCCCTAGA---GGCAGGAGAGTCAATTTCCAGATGCCAGCGTGTCTCC 648
52 ProProGlnGlnGlnGluGluSerLeuArgArgSerGlyArgGlnProLysArgThr 71
649 CTGTCTCCAGCCGCTCTGATAAGACACATGCAAGTCTCTGATAAGAACCCAGGGAGGAG 708
72 GluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgGlnArgSerMetProVal 91
709 AAAGCCGAGAGAGGAGTGTCTGAGAGGAGGAGTTAAACAGTCTCTTATCCACCA 768
92 SerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGluThrAlaSer 111
769 TCACACCGGAG---TGTTGTCCAGCCGTGAGGAGGAGGACCATGAA 813
112 GluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAla 131
814 GAAGCTGTAAAGAAAGAGCTCACAGAACTCTACTTCTGCTCTCCAGATCC--- 870
132 ValLysGluArgProAlaSerSerGluLysValLysGlyGlyAspAspHisAspThr 151
871 -----AAGCCACGACTTGGTGTCTTGGCCAGGGGAGGAGAGATCAAGCCACG 921
152 SerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgGly 171
922 GAGGATAAAAGACAGAA-----AGAGTAAAGAGCCAGAGAG 960
172 -----ArgGlu-GlnGluProThrGluArgProLeuLysGlyLeuGlnSe 186
961 ACCTCCGTGTCCCGCATCTTCAGGCTCCAGCCAGGCTCTGAGGATGATCGTTCAGGA 1020
186 rArgLeuArgLysLysArgGluGluGluGluProAlaGluThrValGlySerGluAlaSe 206
1021 GAGG---CCTCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1077
206 rAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlyValVa 226
1078 GCCCCAGGCGCAATCCTCAGCCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1136
226 lSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIl 246
1137 CCAACCAAGTGTGAAGAGAGAGTGAAGGCTTTGGGGGAGGAGGAGGAGGAGGAGGAGGAG 1196
246 eLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyAspPr 266
1197 AGAGCTGAGTGTCTCCCTGATC-----CCCCACCAAG-TGCTT-----CC 1240
266 oAsnAlaLeuTyCysLeCysArgGlnProHisAsnAsnArg----- 280
1241 TGAAGGCTGTGGTGTATTGGCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1300
281 -----PheMetile-- 283
1301 ATGAGACAGTGTACTCTGATCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1360
284 -CysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaAr 303
1361 CTCTCTGCTACAGTGTCTTCTT-----GAAGTCTGCTGCTATCAGCCAGGAG 1408
303 gGlyArgLeuLeuGluArgAsnGly-GluAspTyIleCysProAsnCysThrIleLeuG 323
1409 AGGA-----CACAGAGGAGGAGGAGAT----- 1431
323 lnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys----- 339
1432 ----GAGGACATGATACAGGATCAGCCGAGGATGAAGAGAGAGTGAAGTCTTCTGCTT 1486
340 -----TrpArgProGly---AspAlaA 346
1487 CCACACCCCTGCAAGTGTCTTCTTCAAGGCTTGGGTGTATCGGCCAGGAGGAGGAGGAG 1546

346 spGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyI 366
1547 AGGAGGAGGAGATGAGGATGTGGATAGT---GAGGATAAGGAGAGATGATTTCAGAGCAG 1603
366 leLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysLysIlePheG 386
1604 CCTTAGGAGAGCTGAGTGCAGACCCACATCCCTCC----- 1638
386 lnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValA 406
1638 ----- 1638
406 laGlnProAspSer-ValTyCysSer-----AsnAspCysIleLeuLysHisAlaAla 423
1639 --CACCCCGAGCAGAGTCCCACTTCAGAGGCTGGGAGATATCGACTCGAAGAGAGAGAG 1696
424 AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMet 443
1697 AGGAAGAGAGAGTGTCTGAGGACTGGGAGAGAGCTGAGCCCTGCCCCCTTCCGAGTGGCA 1756
444 LysMetLysProGluLys-----ProSerLeu 452
1757 TCTATGTACTCGAGAGAGAGCCACCGCTCCCTGGGCTCTCTCTAGGCTGCCCTCCGAC 1816
453 ProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaPro 472
1817 TCAAGAGGAGGCTCAAGCGCCAGAGCCCTACTCATCATCGGAGCCCTGAGACTCCCC 1876
473 GluLysLysGlu-----ThrThrValLysLysAlaValVal----- 485
1877 TAAAGCCAGAGAGGTGGCTTCTCCAGAGAGGTCACTGTCTCATTTCTGCTGTCTGGG 1936
486 -----ProAlaArg-----SerGluAlaLeuGlyLysGluAla 496
1937 CAGGCGCGCCCGCCAGCGCCCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1996
497 AlaCysGluSerSerThrProSerThrProLysAspHisAsnTyAsnAlaValLysPro 516
1997 GCTTCGACAGCGCATCACCC-----AGGCC 2023
517 GluLysThrAlaAlaProSerProSerLeuLeu 527
2024 AGGAGGAGTGAAGCCCTGCTTCACCCCTGCTG 2056
RESULT 15
US-09-105-537-30/c
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30
Alignment Scores:
Pred. No.: 0.00465 Length: 13842
Score: 158.50 Matches: 143
Percent Similarity: 32.62% Conservative: 71
Best Local Similarity: 21.80% Mismatches: 223
Query Match: 5.30% Indels: 221
DB: 3 Gaps: 33

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 15:46:03 ; Search time 3266 Seconds
(without alignments)
5138.563 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPSENEAPKAKPTS.....RSFWTAIPWACRGLGVAALC 562

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frameat_p2n_model -DRV=xlp
-Q/cgn2_1/USPTO.spool_p/US09787016/runat 30042004 083702 4981/app query.fasta_1.711
-DB=EST -Qfmt=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016@cEN 1 4237 @runat 30042004 083702 4981 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gsal:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2976	99.6	2752	11	BC014615 Homo sapi
2	2114.5	70.7	4619	11	AK042474 Mus muscu
3	2100.5	70.3	4651	11	BC060249 Mus muscu
4	1752.5	58.6	1640	11	BC029110 Mus muscu
5	1689	56.5	1614	11	AK014422 Mus muscu
6	1557.5	52.1	1086	9	AL578222 AL578222
7	1403.5	47.0	925	9	AL546121 AL546121
8	1362	45.6	891	13	EX350413 EX350413
9	1300.5	43.5	1018	9	AL550675 AL550675
10	1275.5	42.7	1093	9	AL555810 AL555810
11	1225	41.0	850	14	CD364822 UI-R-FT2-
12	1180.5	39.5	773	12	BG545979 602573278
13	1156	38.7	1033	9	AL580669 AL580669
14	1154.5	38.6	1036	13	BO881380 AGENCOURT
15	1144.5	38.3	760	10	BE298017 601118492
16	1127	37.7	653	10	AW373892 QV0-BT053
17	1127	37.7	653	10	AW373929 QV3-BT053
18	1104.5	37.0	772	10	BE299127 601118879
19	1101	36.8	1201	9	AL514138 AL514138
20	1099	36.8	868	10	BF306721 601888796
21	1089.5	36.5	874	13	BUS22670 AGENCOURT
22	1081	36.2	781	13	BQ921479 AGENCOURT
23	1074	35.9	923	10	BF308603 601890469
24	1043	34.9	1201	9	AL553387 AL553387
25	1041.5	34.8	898	13	BQ646878 AGENCOURT
26	1035	34.6	650	10	BE298593 601119753
27	1027.5	34.4	1119	9	AL550431 AL550431
28	1020	34.1	740	13	BUS89908 UI-CF-FNO
29	1005	33.6	1050	13	EX404270 EX404270
30	1002.5	33.5	3385	11	AK044919 Mus muscu
31	973	32.6	692	10	BE393471 601310142
32	963.5	32.2	653	10	BF308654 601890233
33	948	31.7	660	10	AW960420 EST372491
34	944	31.6	635	13	BQ639998 he22h10.Y
35	942.5	31.5	1201	13	EX442517 EX442517
36	941	31.5	751	10	BF525689 602070017
37	935	31.3	871	13	BUI31981 603116323
38	928.5	31.1	670	13	BU204261 604154862
39	920	30.8	842	13	BQ437271 AGENCOURT
40	919	30.7	510	10	BE293103 601106174
41	910.5	30.5	740	12	B1913500 603179127
42	899	30.1	752	13	BUI11948 603127356
43	890.5	29.8	583	12	BM018416 603646110
44	888.5	29.7	866	13	BQ212019 AGENCOURT
45	864	28.9	512	10	BE245092 TCRAP1D25

ALIGNMENTS

RESULT 1
BC014615
LOCUS
DEFINITION Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift
2752 bp mRNA linear HTC 17-DEC-2003
ACCESSION BC014615
VERSION BC014615.2 GI:33869935
KEYWORDS errors
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2752)

Qy 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1370 AAATCTCAAGATCTTCCAGCTGTGATAGAGCGCGCTGGTCCCTCAAAATGATTGGCCCC 1429
 Qy 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 Db 1430 GGGTGTGTCTCACTGGCGCGAGCCGACTCGGTGTACTGCGAGTAATGACTGTATCTCTCAA 1489
 Qy 421 HisAlaAlaIleThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1490 CAGCCCGCAGCAGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGACAGACCAAGCTTAA 1549
 Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1550 GAAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1609
 Qy 461 IleLysIleSerValHisLysArgProAlaProGluLysGluThrValLys 480
 Db 1610 ATTAATATCTCTTCTGTGCAAGAGACAGACAGCTCCAGAAAAAAGAGACCAAGTAA 1669
 Qy 480 slyeAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSe 500
 Db 1670 GAAGGCGAGTGGTGTCTCTCGCGGAGTGAAGCACTCGGAAGGAGCAAGCTGTGAGAG 1729
 Qy 500 rSerThrProSerTrpAlaSerAspHisAsnTyrAspAlaValLysProGluLysThrAl 520
 Db 1730 CAGCAGCGCTGTGTGGCGAGCGATCACAATTACATGCAAGTAAAGCCAGCAAGACTGC 1789
 Qy 520 alaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAs 540
 Db 1790 TGTCTCTCTCGCTGCTGTGTATAATATGATATCATCTGAGGCTGGCTCTCTGGA 1849
 Qy 540 pProSerArgSerPheTrpIleAlaLeuProTyrAlaCysProGlyLeuGlyValAlaAl 560
 Db 1850 CCCTCCCGTCTTCTTGGATAGCATCCCTCGGCTGTCCAGGACTGGAGTGGAGTGCAGC 1909
 Qy 560 aleuCys 562
 Db 1910 TTGTGT 1916

RESULT 2
 AK042474
 LOCUS
 DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630095A07 product:death inducer-obliterator-1, full insert sequence.
 ACCESSION AK042474
 VERSION 1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tachiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 4619)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, T., Tanaka, F., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. .4619
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM_DB:A630095A07"
 /db_xref="MGI:2407519"
 /db_xref="taxon:10090"
 /clone="A630095A07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 299. .2143
 /note="unnamed protein product; death inducer-obliterator-1 (MGD|MGI:1344352, GB|AJ238332, evidence: BLASTN, 99%, match=2599) putative"
 /codon_start=1
 /protein_id="BAC31270.1"
 /db_xref="GI:2633139"
 /translation="MDDKGLHNSAPKAIKPTSEKPTWGFRTTAKRSCAGDTVDPSSQPPQHHLSLRSGRQPKRTERVEEFITTVRRKKNVPVLSDESPTSTV

FEATURES
 source
 1. .4619
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM_DB:A630095A07"
 /db_xref="MGI:2407519"
 /db_xref="taxon:10090"
 /clone="A630095A07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 299. .2143
 /note="unnamed protein product; death inducer-obliterator-1 (MGD|MGI:1344352, GB|AJ238332, evidence: BLASTN, 99%, match=2599) putative"
 /codon_start=1
 /protein_id="BAC31270.1"
 /db_xref="GI:2633139"
 /translation="MDDKGLHNSAPKAIKPTSEKPTWGFRTTAKRSCAGDTVDPSSQPPQHHLSLRSGRQPKRTERVEEFITTVRRKKNVPVLSDESPTSTV

TDVETASRGSSVESSEIRSPVDSLGKHEPASSEKAKGCEBEEDTSDSDGLTIKE
 LQNRLLRREDEPERSLQNSRLRKKREEDSABTSVQIGSAEQDPUCKQEPER
 SQGPSQSEDDIENQLEKATQNTENPREAGKPECEVDFPNALYICROPHNN
 RFMCCDCEWFHGDYVIGSEARGLIRERNGEDYICNPCTILOQDETSATBQD
 SGRSGVADGDTDSIGTIEQSGEDQIKRIEKAANPSKKKLLKIPYVEAPGAP
 KCIQPGSSVAPDSVSCNDICILKHAATWPLSSGKEQTKPKVKTKPKFSLP
 KCSVOVGKISSEKRLASERENPVKKVLMASRSETSKKAACESSTPWSADHNTN
 AVKPEKPKTALSILKCTYHPKAGPQPSHLGCLGLSTRVLGLVLIVASS
 SLPAKSRITQDASGPVFLPSLWSLSCNFKSCVGLMLBAISYFSFRPW*

ORIGIN

Alignment Scores:

Pred. No.: 1.85e-131 Length: 4619
 Score: 2114.50 Matches: 423
 Percent Similarity: 81.22% Conservative: 31
 Best Local Similarity: 75.67% Mismatches: 95
 Query Match: 70.74% Indels: 11
 DB: 11 Gaps: 7

US-09-787-016A-3 (1-562) x AK042474 (1-4619)

QY 1 MetAspAspGlyAspProSerLeuGluAlaProLysAlaIleValProThrSer 20
 DB 299 ATGGATGATAAAGGCGCCTGAGCATGAGGAGGACCCACAGGCTATCAAAACCCACAGT 358
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaIleArgGluGlyAla 40
 DB 359 AAGGAGTTCAGGAAACCTGGGTTTCGAGACACCGATTGCCAAAGCTGAGGGTGCA 418
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 50
 DB 419 GAGAGACAGGAGTGGACCCAGTGGAGCAGCAACCA-----CAGCAGCATTAACCTCTCC 472
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 473 CTGCGCGCGAGTGACCGGACCAACCAACCTACTGAGAGGGTGAGAGAGTTCTTACCAG 532
 QY 81 AlaArgArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 533 GTTTCGGCGCGAGGAAAGAAAGATGCGCGGTGCTCCCTGAGGATTTCCAGTGGAGGCCCA 592
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 593 TCTTCACAGTCACTAGTGGAGACAGCTTCCGAGGGAGCGGCTGGAAGACAGTCTGAG 552
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 DB 653 ATCAGAAAGTGGCCCTGTATCTGACTCCTTAGGG---AAGAACATCCTGCTCTTCTGAA 709
 QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 710 AAGGCAAAAGGAGTGAAGAGGAGAGACACCTCTGACAGTGCAGTGCATGCGCTTACG 769
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
 DB 770 TTGAAGGAACTTCAGAACCCCTTCGAGAACGAGAGCAGCAACCTGTGGAGAGTCC 829
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200
 DB 830 CTGAGAGGAGTCAAGATCGCTCGAGAAAGACGAGCGCAGAGAGAGATTTCTGCCGAATC 889
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 DB 890 ---GGAGTGTCCAAATAGCAGTCCGAGCAGCAGCAGCAGCTCTCTGTAAAGCAGGAGCT 946
 QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 DB 947 GAGGCTAGTCAGGAGCAGCAGTGTCCAGTCAAGACAGATGACATAGAAATCAGTTGGAA 1006
 QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProProGlyAspLeuGlyArgProLysPro 259
 DB 1007 GGGAGGCGACTCAGGAATACAGAGGAAACCCAGGAGGAGCGGCAACCAACGCT 1066

QY 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
 DB 1067 GAGTGTGAGGTTTACGACCCCAATGCCCCTGTACTGCATCTGCCCCAGCCTTCAACAAC 1126
 QY 280 ArgPheMetIleCysCysAspAspCysGluGluTrpPheHisGlyAspCysValGlyIle 299
 DB 1127 AGGTTTATGATCTGCTGTGATGGGTGTGAGGAGTGGTTCCATGCTGCTGTGGTATT 1186
 QY 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
 DB 1187 TCTCAGGCCCGAGGCGGCTCTCGAAGGAACCGGGAAGACTACATCTGCCCAANTGC 1246
 QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLys 339
 DB 1247 ACCATTTTGCAGTGCAGGATGACACAAACGGTAGGCGCCATGAGCAGACTCTGGG 1306
 QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
 DB 1307 TGCAGATCTGTGGTGTCTGATGGCACACACTGCACAGCATAGGGAACAATAGAGCAG 1366
 QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
 DB 1367 TCCGAGAGAGACCGAGGCATNAAGGTAGGATTGAGAGGCGACCAACCCAGCGGCAAG 1426
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 DB 1427 AAAAACTCAAGATATTCCAGCTGTCTGTAGAGGCTCTGCTGCTCTCTCTCTCTCTCT 1486
 QY 400 ProGlyCysCysHisValAlaGlnProAsnSerValTyrCysSerAsnAspCysIleLeu 419
 DB 1487 CTGTGGTGTTCAGTGTAGACACGCTGACTCTGTGTATTGCTGATGATGATGATGATG 1546
 QY 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
 DB 1547 AAACACGACGACGCTTACCATGAGATTTCTAAGTTTCAAGTAAAGAACAAACAAACCC 1606
 QY 440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
 DB 1607 AAGGAAAAAGTCAAGACGAGGCAAGAAAGTTCAGTCTTCCAAATTCAGTGTTCAGTG 1666
 QY 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVal 479
 DB 1667 GGGATTAAATCTCTTCTGTGCACAGAGACTAGCGCTCAGAGAAAGGAGAAACCCAGTG 1726
 QY 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
 DB 1727 AAGAAA---GTGATGCTGGCTTCCAGGAGTGAAGCTTCTGGGAGGAGGAGCCTGTGAG 1783
 QY 500 SerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThr 519
 DB 1784 AGCAGCACACCATCTCTGGCAAGTGACCACTACAACTGCTGTGAGAGCAGAGAGCA 1843
 QY 520 AlaAlaPro-----SerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyVal 536
 DB 1844 GAGAGGCCCATGCACTCTCGCCACCTTATGATGATAATGTACGTATACCCAAAGGCT 1903
 QY 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
 DB 1904 GGCCTCCCGAGGCCCTCCATC--ATCTGGTGGCT---GCCTGGGGCTGTCTAGGA 1955

RESULT 3

BC060249

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 4651)

BC060249 4651 bp mRNA linear HTC 19-NOV-2003
 Mus musculus death inducer-oblierator 1, mRNA (cDNA clone
 IMAGE:5826043), containing frame-shift errors.

BC060249.1 GI:37805411

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 4651)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Usdin, T.B., Toshiyuki, S., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Loquellano, N.A., Peters, G.J., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Heiton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalski, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 4651)

Strausberg, R.

Direct Submission

Submitted (20-Oct-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-rc@mail.nih.gov

REMARK COMMENT

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada

info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 126 Row: k Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein

This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

1..4651

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:6526043"

/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"

/clone_lib="NIH BMAP_FYO"

/lab_host="DH10B"

/note="Vector: pYX-ASC"

ORIGIN

Alignment Scores:

Pred. No.: 1.64e-130 Length: 4651

Score: 2100.50 Matches: 422

Percent Similarity: 81.07% Conservative: 32

Best Local Similarity: 75.36% Mismatches: 95

Query Match: 70.27% Indels: 12

DB: 11 Gaps: 7

US-09-787-016A-3 (1-562) x BC060249 (1-4651)

QY 1 MetAspAspLysGlyAspProSerAsnGluUaLaProLysAlaLeuLysProThrSer 20

DB 278 ATGGATGATAAAGGGGACCTGAGCAGATGAGGAGGAGCACCAGGCTATCAACCCACG 337

QY 21 LysGluPheArgLysThrTyrGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40

DB 338 AAGGAGTTTACAGGAACCTGGGGTTTCGAGAGACACACGATTCGCCAAGGTGAGGTGCA 397

QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnGlnLeuGlyLeuSer 60

DB 398 GGAGACACGAGGTGAGCACCAGTGTGAGCAGCAACCA-----CAGCAGCATTAACCTCTCC 451

QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80

DB 452 CTGGCCCGCAGTGGACCGGCACCAAAACGATCTGAGAGGTAGAGAGTTTCTTACCAG 511

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100

DB 512 GTTCGGCGCGGAGGGAAGAAATGTCGGGTGCTCTGGAGATTCCTCAGTGCAGCCACA 571

QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120

DB 572 TCTTCCACAGTCAGTGTGAGCAGCTTCGAGGGGAGCGGTGGAAGCAGATTCCTGAG 631

QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140

DB 632 ATCAGAGTGGCCCTGTATCTGCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 688

QY 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspSerGlyLeuThr 160

DB 689 AAGGCAAAAGAGAGGTGAGAGGAGAGAGACCTCTCGACAGTCAGCAGTGTGAGCCTTACG 748

QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180

DB 749 TTGAGGAACTTCAGAACCGCTTCGAGAAAGCAGAGCAGACAGTGTGAGAGGTCC 808

QY 181 LeuLysGlyLysGlnSerArgLeuArgLysArgGluGluGluGluGluGluGluGlu 200

DB 809 CTGAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 868

QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219

DB 869 ---GGAGGTGTCAAATAGCGAGTCCCGGAGCAGACACCTCTCTGTAAAGAGAGCTT 925

QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239

DB 926 GAGGCTAGTTCAGGAGCAGTGTCCAGTCCAGACAGACAGATGACATAGAAAATCAGTTGGA 985

QY 240 GlyLysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysPro 259

DB 986 GGGAGGGCGACTCAGGGAAATACAGAGGAACCCCGAGAGCGGGGCAACCAAGCTT 1045

QY 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279

DB 1046 GAGTGTGAGTTTACGACCCCAATGCCCTGTACTCTCATCTGCGCGAGCTTCACACAC 1105

QY 280 ArgPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyLys 299

DB 1106 AGSTTTATGATGCTGTGTGATCGGTGTGAGGAGTGTGTTCCATGGGTGACGTGTGGGTATT 1165

QY 300 SerGluAlaArgGlyArgLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319

DB 1166 TCTGAGGCCCGAGGCGCGCTCTGGAAGAGACCGGGAGAGACTACATCTGCCCAATTGC 1225

QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLys 339

DB 1226 ACCATTTTGCAGTGCAGATGAGACAAACCGTAGCGCCACCGCAGCAGCAGTCTGGG 1285

QY 340 TtpArgProGlyAspAlaSpGlyThrAspCysThrSerIleGlyThrIleGluGlnlys 359
 Db 1286 TGCAGATCTGGGGTCTGATGGCAGACACTGCACAGCATTAGGACAGTACAGGAG 1345
 QY 360 SerSerGluAspGlnGlyIleGlyGlyArgIleGlyGlyAlaAlaAsnProSerGlylys 379
 Db 1346 TCCGAGAGACCCAGGGGCATTAAGGGTAGGATTGAGAGGCGAGCAAAACCCAGCGGCAAG 1405
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 Db 1406 AAAAACAACCAAGATATTCACGCCCTGTGTAGAGGCTCTGTGTCTTAAATGATGTGC 1465
 QY 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleleu 419
 Db 1466 CTGGGTCTTCAGGTGTAGCAGCGCTGACTCTGTGTATTCAGTATGACTGCTATCTC 1525
 QY 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPr 439
 Db 1526 AAACACGAGCAGCTTACCATGAGTTTCTTAAGTTTCAGTAAAGAACAAACAAACAC 1585
 QY 439 GlyGlyLysMetLysMetLysProGlyLysProSerLeuProLysCysGlyAlaGlnAl 459
 Db 1586 CAAGGAAAGGTCAAGACGACGACGAAAGTTTCAGTTCTTCAAAATGCACTGTTTCAG 1645
 QY 459 aGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluAlaAlaCysG 479
 Db 1646 GGGGATTAATAATCTCTTCTGTGCACAGAGACTAGCTGCAGAGAAAGGAAAGGAAAC 1705
 QY 479 LysLysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysG 499
 Db 1706 GAAGAAA---GGATGCTGGCTTCAGGAGTGAGACTTCTGGAGAGGAGGAGCTGTGGA 1762
 QY 499 uSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysTh 519
 Db 1763 GAGCAGCACACCATCTGGGCAAGTGACCAACATACATCTGTGAAGCCAGAGAGGCC 1822
 QY 519 rAlaAlaPro-----SerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyVa 536
 Db 1823 AGAGAAGGCCCATCTGCCTCCACCTATGTAGTAAATGTAGTATCATCCCAAGGCC 1882
 QY 536 lGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
 Db 1893 TGGCTTCCAGGCGCCCTCCCATC--ATCTGGGTGGCT--GCTGGGGCTGTGTAGGA 1935
 RESULT 4
 BC029110
 LOCUS
 DEFINITION Mus musculus death inducer-obliterator 1, mRNA (cDNA clone
 IMAGE:3980050), containing frame-shift errors.
 ACCESSION BC029110.1 GI:22137689
 VERSION
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1640)
 REFERENCES
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toshibayashi, S.,
 Carrinci, P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McWhan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.C., Blakeley, R.W., Touchman, J.W., Green, E.D.,
 Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: coapbs-femail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
 Series: IRAC Plate: 23 Row: K Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein
 This clone has the following properties: frame shifted.

FEATURES

Location/Qualifiers
 1..1640
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3980050"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI CGAP_Mam1"
 /lab_host="DH10B"
 /notes="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 1,13e-107 Length: 1640
 Score: 1752.50 Matches: 343
 Percent Similarity: 83.83% Conservative: 25
 Best Local Similarity: 78.13% Mismatches: 66
 Query Match: 58.63% Indels: 5
 DB: 11 Gaps: 4
 US-09-787-016A-3 (1-562) x BC029110 (1-1640)
 QY 1 MetaspaspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
 Db 332 ATGGATGATTAAGGGCCACCTGAGCAATGAGAGCAGCCCAAGCTATCAACCCACCACT 391
 QY 21 LysGluPheArgLysThrTyrGlyPheArgThrThrIleAlaLysArgGluGlyVala 40
 Db 392 AAGAGTTTCAGGAAACCTGGGTTTCGAAGAACCAAGATTGCCAACCGTGGGGTGCA 451
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
 Db 452 GGAGACACGGAGGTGGACCCCGAGTGCAGCAGCAACCA-----CAGCAGCATTAACCTCTCC 505
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1086)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:12942093.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1024BD12QP1&cluster=9799.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0D1024BD12QP1.

FEATURES

Location/Qualifiers
1..1086
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1024BD12QP1"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 8.81e-95 Length: 1086
Score: 1557.50 Matches: 292
Percent Similarity: 97.34% Conservative: 1
Best Local Similarity: 97.01% Mismatches: 5
Query Match: 52.11% Indels: 4
DB: 9 Gaps: 2

US-09-787-016A-3 (1-562) x AL578222 (1-1086)

Qy 263 GlyTyrAspProAsnAlaLeuTyrCysIle---CysArgGlnProHisAsnAsnArgPhe 281
Db 1067 GGTACGACCCACGCCC-----TGATGCTTGCSCCGAGCTCACAACAG-FTT 1015
Qy 282 MetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGlu 301
Db 1014 ATGATTTGCTGTGACCGCTGTGAAGAATGTTTCATGCGCATTTGTGGCATTCTGAG 955
Qy 302 AlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIle 321
Db 954 GCTCGAGGAGGCTTTTGGAAAGAAATGGGAAGACTATATCTGCCCAACTGCACCAT 895
Qy 322 LeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArg 341
Db 894 CTGCAGTGGCGATGAGACTCATTCAGAAACGACATCAGCAGGAGCTAATGAGA 835
Qy 342 ProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSer 361
Db 834 CTTGGAGATGCTGATGCGACCGGATTTACAGATAGAGAACTATAGAGCAAGTCTAGC 775
Qy 362 GluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLys 381
Db 774 GAAGACCAAGGATTAAGGGTAGAATTTAGAAAGCTCAATCCAGTGCACAGAGARA 715
Qy 382 LeuIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGly 401
Db 714 CTCAAGATCTTCAGCCTGTGATAGAGCGCCTGGTGCCTCAAAATGTATTGGCCCGGG 655
Qy 402 CysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHis 421

Db 654 TGCTGTCACTGCGCAGCCCGACTCGTCTACTGAGTAATGACTGTATCTCTCAACAC 595
Qy 422 AlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysGlu 441
Db 594 GCGCGACGACCAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCCTAAAGAA 535
Qy 442 LysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIle 461
Db 534 AAGATGAGATGAAGCCAGAGAGCCAGCTCTCCGAATCGGTGCTCAGGCAGGTATT 475
Qy 462 LysIleSerSerValHisLysArgProAlaProGluLysLysGluThrValLysLys 481
Db 474 AAAATCTCTTCTGTGCACAGAGACAGCTCCAGAAAAAAGAGACACAGTGAAGAAG 415
Qy 482 AlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysGluSerSer 501
Db 414 GCAATGCTGCTCTCTGCGGAGTGAAGCACTGGGAGAGAGAGAGAGAGAGAGAG 355
Qy 502 ThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAlaAla 521
Db 354 ACGCGCTGCTGGCGAGCGATCAATTAATGATGATGATGATGATGATGATGATGATGAT 295
Qy 522 ProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAspPro 541
Db 294 CCTCGCGGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 235
Qy 542 SerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAlaLeu 561
Db 234 TCCCGTTCTTCTGATAGCATCCCTCCCTGCGCTGCTGCGAGCTGGGAGTTGCAGCTTTG 175
Qy 562 Cys 562
Db 174 TGT 172

RESULT 7
AL546121 925 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION
AL546121 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CS0D1024Y24 5-PRIME, mRNA sequence.

AL546121
AL546121.2 GI:31267955

EST.

KEYWORDS
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 925)

AUTHORS
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE
Full-length cDNA libraries and normalization

JOURNAL
Unpublished (2001)

COMMENT
On Feb 15, 2001 this sequence version replaced gi:12878929.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 9799.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0D1024BD12QP1&cluster=9799.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID : CS0D1024BD12QP1.

FEATURES

source

1..925

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1024Y24"

/issue type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCWSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 1.65e-84 Length: 925
 Score: 1403.50 Matches: 276
 Percent Similarity: 97.53% Conservative: 0
 Best Local Similarity: 97.53% Mismatches: 7
 Query Match: 46.96% Indels: 5
 DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x AL546121 (1-925)

QY 155 AspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgGlySerGluGln 174
 DB 70 GACAGCGATGCGCTGACCTTGAAGAGCTTCAGATCGCTTCGCGAGGAGCGGAAACAG 129

QY 175 GluProThrGluArgProLeuLysGlyLeuGlnSerArgLeuArgGlySerGluGln 194
 DB 130 GAGCCGACCTGAGGCGCCCTGAAGAGNATCCAGAGTGGCTCGGAAGAGCGCGGAG 189

QY 195 GluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuPro 214
 DB 190 GAGGTCGCCGCGAGACTGTGGCTCCGAGGCGGACACTGTGGAGGCGTCTCTGCC 249

QY 215 SerLysGlnGluProGluAsnAspGlnGlyValValSerGluAlaGlyLysAspArg 234
 DB 250 ACTAAGCAGGAGCGGAGACGATCAGGCGGTTGTGTCCAGGCTGGGAAGATGACAGA 309

QY 235 GluSerLysLeuGluGlyLysAlaGlnAspLysAspGluGluProGlyAspLeu 254
 DB 310 GAGAGTAAGTTGAGGAGAAAGCGCGCTCAGACATCAAGATGAGGAGCTGGAGACTAT 369

QY 255 GlyArgProLysProGluCysGluGlyTyArgProAsnAlaLeuTyCysIleCysArg 274
 DB 370 GCGCGACCGAGGCTGAATGTGAGGCTTACGACCCCAACGCGCTGATTCATTCGCGC 429

QY 275 GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluLysPheHisGly 294
 DB 430 CAGCTTCACAAACAGGTTTATGATTTCTGTGACCGCTGTGAAGATGGTTTCATGCC 489

QY 295 AspCysValGlyLysSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTy 314
 DB 490 GATTGTGTGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGATGGGAGACTAT 549

QY 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
 DB 550 ATCTGCCCAACTGCACCACTTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGCGAGAT 609

QY 335 GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
 DB 610 CAGCAGAGAGCTTAATGGAGACTGAGATGCTGATGGCACCAGTTGTACAGTATAGA 669

QY 355 ThrIleGluGlnLysSerSerGluAspGlnGlyLysGlyArgIleGluLysAlaAla 374
 DB 670 ACAATAGACGAGAGTCTAGCGAAGACCAAGGGATTAAGGGTAGAATTGAGAAAGCTGCA 729

QY 375 AsnProSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 394
 DB 730 ATCCCAAGTGGCAGAGAGAACTCAAGATCTTCAGCCTGTGTATAGAGGCGCTGTGTGC 789

QY 395 SerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyCysSer 414
 DB 790 TCAAAATGATTGGCCCGGGTCTGTACGCTGGCGCA-CCGACCTCGGTGTACTGCACT 848

QY 415 AsnAspCysIleLysLysHisAlaAlaAlaThrMetLysPheLysSerGlyLysGlu 434
 DB 849 NATGACTGTAT-CTCAACACCGCGGAG-ACAATGAAGTTCTA-AGCTCAGGTAAAGA- 904

QY 435 GlnLysPro 437

DB 905 CAGAAGCCA 913

RESULT 8

BX350413/c

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. (Bases 1 to 891)

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 9799.1 For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0BAI036Z802_CS03373_1&cluster=9799.1.

Contact: Peng liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID: CS0BAI036Z802_CS03373_1.

Location/Qualifiers

1. 891

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK002FJ13"

/cell_type="HELA"

/cell_lines="HELA"

/note="1st strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCWSPORT 6 vector. Library was normalized."

FEATURES

source

Alignment Scores:

Pred. No.: 9.77e-82 Length: 891

Score: 1362.00 Matches: 272

Percent Similarity: 93.29% Conservative: 6

Best Local Similarity: 91.28% Mismatches: 19

Query Match: 45.57% Indels: 4

DB: 13 Gaps: 1

US-09-787-016A-3 (1-562) x BX350413 (1-891)

QY 24 ArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAlaGlyAspAla 43

DB 890 AGGAAACCA---GGGTTTTCGAGAGGCCACTATCGCCAG-CAGAGGGGCCCGAGGACCG 835

QY 44 GluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSerLeuArgArg 63

DB 834 NAGNCTGACCCACTGAGCGCGCACCCACAGCAGCAGCTGGGCTGTCTCTNCGGCGC 775

QY 64 SerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIleAlaArg 83

DB 774 AGTGAAGACCCCGCCAGCGACTGAGCGCGTGGAGAGTTCCTCAGCAATCGCGGCGC 715

QY 84 ArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThrSerCysPro 103

DB 714 GCGGCAAGNAGACATGCGCTGTCTCCCTGAGGATTCCTGAGCCACACCTCTGCGCC 655

QY 104 AlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSer 123

```
Db 654 GCCACAGACGCCGAGACAGCTCGAGGGGCGAGCGTGGAAAGCGCTTCTGAGACCGACGAGC 595
Qy 124 GlyProGlnSerAlaSerThrAlaValLysGluValArgProAlaSerSerGluLysValLys 143
Db 594 GGCCTCCAGCTGCTGCTCCAGCTGTGAAGGACGACCGAGCGCTTCTGAAAGGAGGAAA 535
Qy 144 GlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGlu 163
Db 534 GAGGGGATGACACCATGACACCTCCGATAGTGTGACAGCGATGGCTGACCTTGAAGAG 475
Qy 164 LeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgProLeuLysGly 183
Db 474 CTTTCAGATCGCTTCGCGAGGAGCGGAGACGAGGCGCCACTGAGAGCGCCCTGAAAGGG 415
Qy 184 IleGlnSerArgLeuArgLysLysArgGluGluGluGluGluGluGluGluGluGluGlu 203
Db 414 ATCCAGAGTGGCTGCGGAGAGGCGCCGGGAGGAGGCTCCCGCCGAGAGTGTGGGCTCC 355
Qy 204 GluAlaSerAspThrValGluGluValLeuProSerLysGlnGluProGluAsnAspGln 223
Db 354 GAGGCCAGTGACACTGTGAGGCGGCTGCTGCCAGTAAAGCAGGAGCCCGAGAACGATCAG 295
Qy 224 GlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGluGluGluGlu 243
Db 294 GGGGTTGTGCTCCAGGCTGGAAAGATGACAGAGAGAGTAAAGTTGGAGGAGAGGCGGCT 235
Qy 244 GlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGly 263
Db 234 CAGGACATCAACATGAGGAGCGCTGAGAGCTTGGGCGGACCGAGCGCTGATGTGAGGCT 175
Qy 264 TyrAspProAsnAlaLeuThrCysArgGluProHisAsnArgPheMetIle 283
Db 174 TAGGCCCCAACCCCTGTATTTGATTTGCGGCGAGCCTCACAAACAGAGTTATGATT 115
Qy 284 CysCysAspArgCysGluGluTppHeHisGlyAspCysValGlyIleSerGluAlaArg 303
Db 114 TGCTGTGACCGCTGTGAAGATGCTTTCATGGCGATGTGTGGGCTTCTGAGGCTCGA 55
Qy 304 GlyArgLeuLeuGluArgAsnGlyGluAspThrIleCysProAsnCysThrIle 321
Db 54 GGGAGGCT-TTGGAAAGGAATGGAA-GACTATATCTGCCCAACTGCACCAT 3
```

RESULT 9 9
AL550675/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL550675 1018 bp mRNA linear EST 31-MAY-2003
AL550675 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
Clone CS0D1056YK04 3-PRIME, mRNA sequence.
AL550675
AL550675.2 GI:31272492
EST.
Homo sapiens (human)
Homo sapiens
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1018)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12887874.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1056BF02NP1&cluster=9799.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1500
Paradise Avenue Genoscope sequence ID : CS0D1056BF02NP1.
Location/Qualifiers
1. 1018

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1056YK04"
/tissue="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores:
Pred. No.: 1.59e-77 Length: 1018
Score: 1300.50 Matches: 249
Percent Similarity: 93.96% Conservative: 0
Best Local Similarity: 93.96% Mismatches: 15
Query Match: 43.51% Indels: 2
DB: 9 Gaps: 1
US-09-787-016A-3 (1-562) x AL550675 (1-1018)

Qy 298 GlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysPro 317
Db 908 GGCATTTCTGAGCTCGAGGGA---GCTTTTGAAGGAATGGGA-GACTATATCTGCCCC 853
Qy 318 AsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlu 337
Db 852 AACTGCACCATTTCTCAAGTGCAGGATGAGACTCATTCAGANACCGCAGATCAGCAGGAA 793
Qy 338 AlaLysTroArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlu 357
Db 792 GCTAATGAGACCTTGGAGATGCTGATGACCGATTTGACATATAGACATAGAG 733
Qy 358 GlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSer 377
Db 732 CAGAGTCTMGCGAGAGACCAAGCGATTAAGCGTGAATTTGAGAAAGCTGCAATCCAACT 673
Qy 378 GlyLysLysLysLysLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCys 397
Db 672 GGCAGAGATGATCTCAAGATCTTCAGCGCTGTGATAGAGCGCGCTGGTGCCTCAAAATGT 613
Qy 398 IleGlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCys 417
Db 612 ATTGCCCCCGGTGCTGTCACTGCGGCGAGCCCGACTCGTGTGACTGCAATGACTGT 553
Qy 418 IleLeuLysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysPro 437
Db 552 ATCTCTCAAACACCGCGCAGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGAGCCCA 493
Qy 438 LysProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyVala 457
Db 492 AAGCCTTAAGMAAAGATGAGATGAGCCAGAGAGCCCGACTTCCGAAATGCGTGTCT 433
Qy 458 GlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThr 477
Db 432 CAGCGCAGGTATTAAATCTCTTCTGTGTGCAACAGAGACCCAGCTCCAGAAAAAAGAGACC 373
Qy 478 ThrValLysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAla 497
Db 372 ACAGTGAGAGAGAGTGTGTGCTCCCTGCGGAGTGAAGCACTCCGAGAGAGAGTGTCT 313
Qy 498 CysGluSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGlu 517
Db 312 TGTGAGAGCAGCAGCGCGCTGCTGGGCGAGCGATCAATTAATGACATTAAGCCAGAA 253
Qy 518 LysThrAlaAlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGly 537
Db 252 AAGACTGTGCTCCCTCGCGCTGACTGTGTATTAATGATGATATACCTTAGGGGTGGC 193
Qy 538 LeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGly 557
Db 192 CTCCTGGAGCCCTCCCGTCTTCTGTGATAGCATCCCTCGGCGCTGTCCAGACTGGGA 133


```

/db_xref="taxon:9606"
/clone="UI-H-F72-bjm-d-20-0-UI"
/tissue_type="Aveolar Macrophage"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP F72"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP F72 is a subcloned cDNA library constructed from
a pool of 81 RNA samples from Alveolar Macrophages
challenged with different treatments. The library was
subcloned according to Bonaldo, Lennon and Soares, Genome
Research, 6:791-806, 1996. The tissue was provided by Dr.
Gary W. Hunninghake of the University of Iowa.
TAG RTSSUS=Human Lung Aveolar Macrophage
TAG_LIB=UI-H-F72
TAG_SEQ=GGCCATGCGG"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1-52e-72 Length: 850
Score: 1225.00 Matches: 229
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 2
Query Match: 40.98% Indels: 0
DB: 14 Gaps: 0

US-09-787-016A-3 (1-562) x CD364822 (1-850)

QY 332 ThrAlaaspGlnGluAlaLysTyrAtpProGlyAspAlaAspGlyThrAspCysThr 351
DB 850 ACGGAGATCAGCAGGAAGCTAAATGAGACCTCGAGATCTGATGCGCCGATTTGACA 791
QY 352 SerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGlu 371
DB 790 AGTATAGACATAGCAGCAGAGTCTAGCGAAGACCAAGCGATTAAGGCTAGATTGAG 731
QY 372 LysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 391
DB 730 AAAGCTGCANATCCAGTGCNAAAGAACTCAAGATCTTCCAGCTGTGTAGAGGCG 671
QY 392 ProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerVal 411
DB 670 CTTGGTCCCTCAAAATGTTTGGCCCGGTGCTGTGTCAGTGGCGCAGCCGAGCTGGTG 611
QY 412 TyrCysSerAsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSer 431
DB 610 TACTGCAGTAATGACTGATCTCTCAACACGCGCCGACGACATGAAGTTTCTAAGCTCA 551
QY 432 GlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSer 451
DB 550 GGTAAAGACAGAGCCCAAGCCCTAAAGAAAGATGAAGTGAAGCAGAGAGCCGAGT 491
QY 452 LeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysAsgProAla 471
DB 490 CTTCCGAAATGCGGTGCTCAGGCAGGTATTAAATCTCTTCTGTGCACAAGAGACCAAGCT 431
QY 472 ProGluLysLysGluThrThrValLysLysAlaValValProAlaAsgSerGluAla 491
DB 430 CCGAATAAAGAGACACACAGTGAAGAGCAGTGTGTGTCCTCGCGAGGTGAAGCA 371
QY 492 LeuGlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerAspHisAsnTyr 511
DB 370 CTCGGAGAGGAGCAGCTTGTGAGCAGCAGCAGCCGTCGTGGGCGGAGCATCAATTC 311
QY 512 AsnAlaValLysProGluLysThrAlaAlaProSerProSerLeuLysCysMet 531
DB 310 AATGACGTAAGACCAAGAAAGACTGCTGCTCCCTCGCGGTCTACCTGTGTATTAATGTATG 251
QY 532 TyrHisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTyr 551
DB 250 TATCACCTAGGGGTGGCTCTCTGAGACCCCTCCGCTCTTCTGATAGCATCCCTCGG 191

```

```

QY 552 AlaCysProGlyLeuGlyValAlaAlaLeuCys 562
DB 190 GCCTGTCCAGGACTGGAGATTGCAGCTTTGTGT 158

```

RESULT 12

```

BG545979
LOCUS
DEFINITION
602573278F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4701475 5',
mRNA sequence.

```

```

ACCESSION
BG545979
VERSION
BG545979.1 GI:13544644
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 773)

NIH-MGC <http://mgc.nci.nih.gov/>;
National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLM1536 row: 1 column: 20

High quality sequence stop: 699.

FEATURES

source

1..773

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4701475"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_77"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CAGCGCCATTATGCCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1-33e-69 Length: 773

Score: 1180.50 Matches: 234

Percent Similarity: 92.28% Conservative: 5

Best Local Similarity: 90.35% Mismatches: 12

Query Match: 39.49% Indels: 8

DB: 12 Gaps: 1

US-09-787-016A-3 (1-562) x BG545979 (1-773)

QY 238 LeuGluGlyLysAlaAlaGlnAspIleLysAspGluGlnProGlyAspLeuGlyArgPro 257

DB 2 TTGAGGGAAGAGCGCGCTCAGGACATCAAGATGAGAGCGCTTGAGACTTGGCCGACCG 61

QY 258 LysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHis 277

DB 62 AAGCCTGAATGTGAGGGTTAGACCCACAGCCCTGTATTGCAATTCGCCAGCCTCAC 121

QY 278 AsnAsnArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysVal 297

DB 122 AACACAGGTTTATGATTTCTCTGTGACCGCTGTGAAGAATGGTTTCTATGCGGATTGTGTG 181

QY 298 GlyIleSerGluAlaArgGlyValGluLeuLeuGluArgAsnGlyGluAspThrIleCysPro 317
Db 182 GGCATTTCGAGGCTCGAGGGAGGCTTTTGGAAAGCAATGGGAGACTATATCTGCCCA 241
QY 318 AsnCysThrIleLeuGluValGlnAspGluThrHisSerGluThrAlaAspGlnGlu 337
Db 242 AACTGCACATCTCTCAAGTGCAGATGAGACTATTGAAACCGCGAGATCAGCAGAA 301
QY 338 AlatyStpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlu 357
Db 302 GCTAAATGGAGACCTGGAGATCTGATGCGCCGATTGTACAAAGTATAGGAACAATAGAG 361
QY 358 GlnIleSerSerGluAspGlnGlyIle-LysGlyValGlnGluAlaAlaAsnProSe 377
Db 362 CAGAAAGTCTAGCGAAGACCAAGGATACAGGGTAGAATTGAAAGCTGCANATCCAG 421
QY 377 rGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerIleCys 397
Db 422 TGGCAAGAAGAACTCAAGATCTTCAGCCTGTGATAGAGGCGCTGTGCTCAAAATG 481
QY 397 sIleGlyProGlyCysCysHisValalacInProAspSerValTyrCysSerAsnAspC 417
Db 482 TATTGGCCCCGGTGTGTGACGTGGCGAGCCGACTCGGTGTACTCGATTAATGATG 541
QY 417 sIleLeuLysHisAlaAla-AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysP 437
Db 542 TATCTCAACACGCGCGCAGGACATGAAGTTTCTTAAGCTCAGTAAAGAACAGAAAGC 601
QY 437 rGlyProLysGluLysMet-LysMetLysProGlu-LysProSerIleProLysCys 455
Db 602 CAAAGCCTTACAGAAACGATGAGATGAGTGAAGCCAGAGAAACCCAGTCTTCGAAATGC 661
QY 456 GlyAla-GlnAlaGlyIle-LysIleSerSerValHisLysArgProAlaProGluLysL 475
Db 662 GGGCTTCAGGCGGGTATTCACCACTCTCTCTGTGCAACAAGAGACCACTCCAGAAAG 721
QY 475 ysGluThrThr-ValLysLysAla-ValValProAlaArg 488
Db 722 AAGAGAACACAGGTGAAGAAGCCAGGGGTGTCTCCGTGGCGC 764

RESULT 13

AL580669 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DJ007YD06 5-PRIME, mRNA sequence.
AL580669
VERSION 1
KEYWORDS Full-length cDNA libraries and normalization
SOURCE Unpublished (2001)
ORGANISM On Feb 16, 2001 this sequence version replaced gi:12946913.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ007YD063Qp1&cluster=9799.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DJ007YD03Qp1.

Location/Qualifiers
1. .1033
/organism="Homo sapiens"
/mol_type="mRNA"

FEATURES

source

1033

1028

1023

1022

1021

1020

1019

1018

1017

1016

1015

1014

1013

1012

1011

1010

1009

1008

1007

1006

1005

1004

1003

1002

1001

1000

999

998

997

996

995

994

993

992

/db xref="taxon.9606"
/clone="CS0DJ007YD06"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
Pred. No.: 8,57e-68 Length: 1033
Score: 1156.00 Matches: 231
Percent Similarity: 96.69% Conservative: 3
Best Local Similarity: 95.45% Mismatches: 8
Query Match: 38.68% Indels: 1
DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x AL580669 (1-1033)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 304 ATGGACGACAAAGCGGACCCCGAGGANTGAGGAGCCTTAAGGCATCAAAACCCACGAGC 363
QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 364 AARAGTTTCAGAAATCATGGGGTTTTCGAGGAGCACCATATCGCCAAAGCGAGGCGCGCA 423
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 424 GGGACGCGGAGGTGACCCACTGGAGCGCCGACCCACACAGCAGCAGCTGGGCTGTCC 483
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 484 CTGGGCGCAGTGGGAGGAGCGCCCAAGCGCACTGAGCGCGTGGAGCAGTCTCTGACCAT 543
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 544 GCGGCGCGCGCGGAGGAGCATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 603
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 604 TCTGCTCCCGCCACAGACGCGCGAGCAGCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAG 663
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 664 ACCAGAGCGGCGCGCGGAGTCTGCTTCCACAGCTGTGAAGGAAACGACCGCTCTTCTGAA 723
QY 141 LysValLysGlyValAspAspHisAspAspThrSerAspSerAspSerAspSerLeuThr 160
Db 724 AAGGTGAAGGAGGAGGAGTGAACAAGATGACCTCCGATAGTCAGCAGCGATGGCTGACC 783
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Db 784 TTGAARAGCTTCAGAAATCGGCTTCGAGGAAGCGGGAACAGGAGCCACCTGAGAGGCGCC 843
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysValArgGluGluGlyProAlaGluThr 200
Db 844 CTGAAGGAATCCAGAGTCCGCTGCGGAARAGCGCGGAGGAGGCTCCCGCGAGACT 903
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 904 GTGGCTCCGAGGCGCARTGACATATGAGGGGCGCTCTKSCCART-AARCAAGAGCGCCG 962
QY 221 AsnArgGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGly 240
Db 963 AACGATCAAGGGGTGTGTCTCCAGGCTGGGAAAGAAAGAGAGAGAAATAGTTGAGGGA 1022
QY 241 LysAla 242
Db 1023 AAGGCG 1028

Db 190 GATGACAGAGAGTAAGTTGGAGGGAAGGGCGCTCAGGACATCAAGATGAGGAGCCT 249

QY 252 GlyAspLeuGlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCys 271

Db 250 CGAGAGCTTGGGCGGACCGAAGCCTGAATGTGAGGGTTACGACCCCAACCCCTTGA-TGT 308

QY 272 IleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluTyr 291

Db 309 TTTTGGCCGAGCCTCACCAACAGCTTATGATTTGCTGTGACCGCTGTGAAGATGG 368

QY 292 PheHisGlyAspCysValGlyIleSerGluAlaAa-gglyArgLeuLeuGluArgAsnGly 311

Db 369 TTTTCATGGCGATTGTGTGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGATGG 428

QY 312 GluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGlu 331

Db 429 GAAGACTATATCTGCCCAACTGCACCATCTGCAAGTCGACAGTGAAGTACTTATTCAGA 488

QY 332 ThrAlaAspGlnGlnAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThr 351

Db 489 ACAGGAGATCAGCAGGAGCTAAATGGAGACTTGGAGATGCTGATGGACCCGATTTGACA 548

QY 352 SerIleGlyThrIleGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGlu 371

Db 549 AGTATAGGAACTAATAGCAGAGTCTAGCAAGACCAAGGAGTAAAGGGTGAAGTGG 608

QY 372 LysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLysLysLysLysLys 391

Db 609 ATAGCTGCAATCCAGTGGGCAAGAACTCAAGATCTTCCAGCTTGTGTAGAGGC 668

QY 391 aProGlyAlaSerLysCysIleGlyPro-----GlyCysCysHisValAlaGlnProAs 409

Db 669 GCCTGGTGGCTCAAAATGTATTGGCCCCCGGGTGGCTGTCCACGCGGGCGCCGACCCCA 728

QY 409 pSerValTyrCysSerAsnAsp-----CysIleLeuLysHisAlaAlaAlaThrMe 426

Db 729 C-----CTCGGGTGTACTGCAAGTAAATGACTTGGATCCCTCAAAACCGCCGCGCA 785

QY 426 tLysPheLeuSerSer-----GlyLysGluGlnLysProLysProLysGlu 441

Db 786 CAAATGAAGTCTCTTAAGCCTTCGGGTAAAGGAACACGCAAGCCCGCCGCGCTTAAA 845

QY 441 uLysMetLysMetLysProGluLysProSer-----LeuProLysCysGlu 456

Db 846 GAAAGAGTGAAGTATTTGAAGCCCGGAGAGGCCCCCAGAGGCCCCCTTCCCTTAATGC-- 903

QY 456 yAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysGlu 476

Db 904 -----GCGGGTGCTCCCGGCGCAAGCGTATTAAATCCTCTTCTGTGGGCGCCCA 956

QY 476 uThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAl 496

Db 957 ACAAGACCCCGCTCTCTCGAAGATAAACCCGACGACCCCGCCGCGGACACACAGCGTAG 1016

QY 496 aAlaCysGluSerSer 501

Db 1017 TCGGGCGGGGTCTTCC 1032

RESULT 15

BE298017

LOCUS 601118492F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3028271 5',

DEFINITION mRNA sequence.

ACCESSION BE298017

VERSION BE298017.1 GI:9181603

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 760)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

RESULT 14

BQ881380

LOCUS AGENCOURT_8725620 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6340359

DEFINITION s' mRNA sequence.

ACCESSION BQ881380

VERSION BQ881380.1 GI:22273398

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1036)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-t@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LICM2536 row: k column: 16
High quality sequence start: 41
High quality sequence stop: 312.

FEATURES

source

1. 1036

Location/Qualifiers

/organism="Homo sapiens"

/mol type="mRNA"

/db xref="taxon:9606"

/clone="IMAGE:6340359"

/tissue_type="neuroblastoma, cell line"

/lab host="DH10B (phage-resistant)"

/clone lib="NIH_MGC_47"

/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1.09e-67 Length: 1036

Score: 1154.50 Matches: 249

Percent Similarity: 75.14% Conservative: 11

Best Local Similarity: 71.97% Mismatches: 66

Query Match: 38.62% Indels: 22

DB: 13 Gaps: 6

US-09-787-016A-3 (1-562) x BQ881380 (1-1036)

QY 172 ArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLys 191

Db 11 CGGTCTACTCGGGACTGAGAGGCCCTTACAGA-ATACAGAGTCCCTGCGGAGAAG 69

QY 192 ArgArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGly 211

Db 70 CGCGGGAGGAGGGTCCCGCGAGACTGTGGCTCCGAGGCCAGTGCACACTGTGGAGGGC 129

QY 212 ValLeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLys 231

Db 130 GTCTGCCCCAGTACAGAGGAGCCCGAGACATCAGGGGGTGTGTGCCAGGCTGGGAA 189

QY 232 AspAspArgGluSerLysLeuGluGlyValAlaGlnAspThrLysAspGluPro 251

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLM86 row: g column: 24
High quality sequence stop: 610.
Location/Qualifiers
source
1..760
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3028271"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 17"
/note="Organ: muscle; vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.:	3,44e-67	Length:	760
Score:	1144.50	Matches:	224
Percent Similarity:	93.85%	Conservative:	5
Best Local Similarity:	91.80%	Mismatches:	13
Query Match:	38.29%	Indels:	3
DB:	10	Gaps:	1

US-09-787-016A-3 (1-562) x BE298017 (1-760)

QY	319	CysThrIleuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGluAla	338
DB	3	TGCACCAFTCTGCAAGTCGAGGATGAGACTCATTGAGAACCGCAGATCAGCGAGAGCT	52
QY	339	LysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGln	358
DB	63	AAATGGAGACCTGGAGATGCTGATGCAACGATTGTACAGTATAGCAACAATAGAGCAG	122
QY	359	LysSerSerGluAspGlnGlyIleGlyGlyValIleGluValAlaAlaAspProSerGly	378
DB	123	AACTCTAGCGAGACCAAGCGATAGAGGGTAGAATTGAGAACTGCAATCCAGTGGC	182
QY	379	LysLysLysLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIle	398
DB	183	AGAGAGAACTCAAGATCTTCCAGCTGTGATAGAGCGCTGGTGCCTCAAAATGTATT	242
QY	399	GlyProGlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIle	418
DB	243	GGCCCCGGGTGCTGTCACGTGGCGCAGCCGACTCGGTGACTGCTGATGCTATC	302
QY	419	LeuLysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys	438
DB	303	CTCAACACCCCGCAGCGCAATGAAGTTTCTTAAGCTCAGGTAAGAACAGAGCCAAAG	362
QY	439	ProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGln	458
DB	363	CCTAAAGAAAGATGAGATGAAGCCAGAGAGCCAGTCTCCGAAATGCGGTGCTCAG	422
QY	459	AlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThr	478
DB	423	GCAGGTATTAAATCTCTTCTGTGCACAGAGACCAGCTCCAGAAAAAGAGACCACA	482
QY	479	ValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCys	498

Db	483	GTGAAGAGGAGGAGTGTGTGTCCTTCGCCGAGTGAAGCACTCGGGAAGGAGCAGCTTGT	542
QY	499	GluSerSerThrProSerTrpAlaSerAspHisAsnTyzAsnAlaValLysProGluLys	518
DB	543	GAGAGCAGCAGCGCGTCTGTGGGCGAGCGATCAAAATTACAAATGCAATGAGCCAGAAAA	602
QY	519	ThrAlaAlaProSerProSerLeuLeuTyLysCysMetTyHisLeuGlyValGlyLeu	538
DB	603	GACTGTGCTCCTCGCCGTCACCTGTTCTTAATGTGTGTTCACCTAGGGGTGGGCTC	662
QY	539	LeuAspProSerArgSerPheTrpIleAlaIleProTrpAla-CysProGlyLeuGlyVa	558
DB	663	CTGGACCCCTC-CGTTCTTTCGGGATGGCT---CCCTGGGCTGTGACGGCGGGTGGTTG	718
QY	558	LalaAlaLeu	561
DB	719	TGTAGCGATC	728

Search completed: May 1, 2004, 18:40:04
Job time : 3324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: May 1, 2004, 14:32:06 ; Search time 554 Seconds
(without alignments)
4309.544 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPSEHAPKAIKPTS.....RSFWIAIPNACPLGVAALC 562

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q/cgn2_1/USFIO_spool_p/US09787016/runat 30042004 083701 4961/app query fasta_1.711
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1 1 708 @runat 30042004 083701 4961 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2989	100.0	2610	3	Aaz90578 Human dea
2	2981	99.7	2772	4	Aah15102 Human cdn
3	2742	91.7	2085	5	Abv28029 Human pro
4	2742	91.7	2085	5	Abv27935 Human pro
5	2742	91.7	2085	5	Abv22095 Human pro
6	2737	91.6	2085	5	Abv22192 Human pro
7	2449	81.9	7838	4	Aak52368 Human pol
8	2440	81.6	7290	4	Aak53352 Human pol

9	2262	75.7	2332	5	Aas86576 DNA encod
10	2112.5	70.7	2867	3	Aaz90579 Murine de
11	1706	57.1	1764	3	Aac77933 Human can
12	1467	49.1	2986	5	Aas86575 DNA encod
13	989	33.1	30626	4	Aak67051 Human imm
14	898	30.0	503	5	Aas86574 DNA encod
15	807	27.0	478	8	ACH31915 Human end
16	785	26.3	481	5	ABV42461 Human pro
17	561.5	18.8	362	5	ABV12528 Human pro
18	549	18.4	681	5	Aah08005 Human cdn
19	527	17.6	365	5	ABV42575 Human pro
20	527	17.6	365	5	ABV33664 Human pro
21	494	16.5	324	6	ABN21248 Human ORF
22	471	15.8	408	3	AAA43780 Mouse sec
23	447	15.0	7838	4	Aak52368 Human pol
24	420	14.1	256	5	ABV03359 Human pro
25	329.5	11.0	6051	4	Ab129341 Drosophil
26	311	10.4	7174	9	ADC37602 Human nuc
27	267.5	8.9	8751	4	ABL29340 Drosophil
28	246	8.2	636	5	ABV03229 Human pro
29	232.5	7.8	2387	3	AAF22374 Human sec
30	232	7.8	5975	6	ABL61775 Colon ade
31	232	7.8	6256	4	Aai58390 Human pol
32	232	7.8	6256	8	ADB48359 Novel hum
33	232	7.8	6948	5	Aah81806 Human dif
34	232	7.8	7354	4	Aak52041 Human pol
35	226	7.6	6584	4	Aai60176 Human pol
36	226	7.6	6996	4	Aak53025 Human pol
37	221.5	7.4	6947	7	ACF34533 Gene enco
38	219.5	7.3	593	5	ABV12398 Human pro
39	217	7.3	1992	4	ABL27413 Drosophil
40	214.5	7.2	2320	5	ABX71405 Human tes
41	212.5	7.1	1161	4	ABL27411 Drosophil
42	203.5	6.8	10314	4	ABL29797 Drosophil
43	203	6.8	659	3	Aac08973 Human sec
44	203	6.8	915	4	Aas26292 Human cdn
45	203	6.8	915	7	ABX73633 Human nov

ALIGNMENTS

RESULT 1
Aaz90578

ID Aaz90578 standard; cdna; 2610 BP.

XX AC Aaz90578;

XX DT 19-JUN-2000 (first entry)

XX DE Human death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.

XX KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antirheumatic; antiinflammatory; antiproliferative; ss.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

XX CDS 265..1953

XX FT /*tag= a

XX FT /product= "DIO-1"

XX PN WC2000015787-AL.

XX PD 23-MAR-2000.

XX PF 10-SEP-1999; 99WO-GB003019.

XX PR 10-SEP-1998; 98SE-00003069.

XX PR 17-SEP-1998; 98US-0100873P.

XX PA (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.

PA (BANN/) BANNERMAN D. G.

XX Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
 XX WPI; 2000-271426/23.
 DR P-PsDB; AAY67579.
 XX
 PT New DNA encoding human and murine death inducer-obliterators 1
 PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
 PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
 PT hyperproliferative skin disorders.
 XX
 PS Claim 1; Fig 1A; 27pp; English.
 XX
 CC The invention provides nucleic acids encoding the human and murine death
 CC inducer-obliterators 1 (DIO-1) polypeptides. The polypeptides can be
 CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
 CC agonists and antagonists are used as a medicament for treating diseases
 CC characterized by an alteration in cell death or by hyperproliferation,
 CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
 CC tumours, malignant tumours or hyperproliferative skin disorders. They are
 CC also useful in the treatment of metabolic, proliferative or inflammatory
 CC conditions. The present sequence represents a cDNA encoding the human DIO
 CC -1 polypeptide
 XX
 SQ Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,73e-177 Length: 2610
 Score: 2989.00 Matches: 562
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-09-787-016a-3 (1-562) x AAZ90578 (1-2610)

QY 1 MetAspLysGlyValSerProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 DB 265 ATGGACGACCAAGGCGACCGCGACATGAGGCGACCTAAGGCGACCAACCCACCGC 324
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrIleAlaLysArgGluGlyAla 40
 DB 325 AAAGAGTTTCAGGAAACATGGGGTTTTTCGAAGGACCACTATCGCCAGCGAGGCGCGCA 384
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 DB 385 GGGGACCGGAGGCTGACCCACATGGAGCGCCGCCACCCCCACAGCGAGCTGGGCTGTCC 444
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 445 CTGGCGCGCAGTGGGAGCGAGCCCAAGCGCACTGAGCGGCTGGAGCAGTTCTCGACCAAT 504
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 505 GCGCGCGCGCGCGCGAGGAGGAGTCCTGTCTCCCTGGAGGATTTCTGGTGGAGCCCGC 564
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 565 TCCTGCGCCGCCACAGACGCGGAGACAGCGCTCCGAGGCGAGCGTGGAAAGCGCTCTGAG 624
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
 DB 625 ACCAGAGCGCGCCCGCAGTCTGTTCTCCACAGCTGTGAAGGAAAGACCGAGCTTCTGAA 684
 QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 685 AAGGTGAAGAGGAGGATGACCGATGACACCTCCGATAGTCACAGCGATGGCTGACC 744
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 DB 745 TTGAAGAGCTTCAGATCCCTTCGAGAGCGGAAACAGAGGCGCCACTGAGAGGCCCC 804
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluGlyProAlaGluThr 200

DB 805 CTGAAGAGGATCCAGAGTGCCTCGCGAAGAAAGCGCGGAGGAGGTCCTCCGCGAGACT 864
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 DB 865 GTGGCTCCGAGCGCAGTGCACACTGTGGAGGGGCTCTGCCAGTAAGCAGGAGCCCGAG 924
 QY 221 AsnAspGlnGlyValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 DB 925 AACGATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATCAGAGAGAGTAAGTTGGAGGA 984
 QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGluAspLeuGlyArgProLysProGlu 260
 DB 985 AAGCGGCTCAGGACATCAAGATGAGAGCTGGAGACTTGGCGCGAGAGCCCTGAA 1044
 QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnArg 280
 DB 1045 TGTGAGGGTTAAGACCCCAAGCGCTGTATTGCAATTTGCCGCCAGCTCTCAACAACAGG 1104
 QY 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSer 300
 DB 1105 TTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTTCATGGCGATTTGTGGGCAATTCT 1164
 QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 DB 1165 GAGCTCGAGGGAGGCTTTTGGAAAGGAATGGGAGACTATATCTGCCCAACTGCACC 1224
 QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
 DB 1225 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGCTTAATGG 1284
 QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 DB 1285 AGACTGGAGATGCTGTGGCACCGATTGTACCAAGTATAGGAACAATAGAGCAGAGTCT 1344
 QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 DB 1345 AGCGAAGACCAAGGATAAAGGGTAGAATTGAGAAGCTGCAAAATCCAGTGGCAGAGAG 1404
 QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 DB 1405 AAACCTCAGATCTTCCAGCTGTGTAGAGGCGCTGTCTCAAAATGATTATGGCCCC 1464
 QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLysLys 420
 DB 1465 GGGTCTCTGACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCTCTCAA 1524
 QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 DB 1525 CACGCGCGAGCGACCAATGAGTTTCTAGCTCAGGTAAAGAACAGAACCCAGGCTTAA 1584
 QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnLysGly 460
 DB 1585 GAAAGATGAAGATCAGCGAGAGAGCCAGTCTTCCGAAATGCGGTGTCTCAGGCGGT 1644
 QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 DB 1645 APTAAATCTCTTCTGTGCAAGAGACAGCTCCGAAAGGAGAGAGAGCAGCAGTGAAG 1704
 QY 481 LysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 DB 1705 AAGGCAAGTGGTGGTCTCCCTCGCGAGTAGAGCACTCCGGAAGAGAGAGCTTGTGAG 1764
 QY 501 SerThrProSerTrpAlaSerAspHisSerThrAlaValLysProGluLysThrAla 520
 DB 1765 AGCAGCGCGTGTGGGCGAGCGATCAATTAATGATGATGATGATGATGATGATGATGAT 1824
 QY 521 AlaProSerProSerLeuLysCysMetTyrHisLeuGlyValGlyLeuLysAsp 540
 DB 1825 GCTCCCTCGCGCTCACTGTGTATTAATGATGATGATGATGATGATGATGATGATGAT 1884
 QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560

Db 1885 CCCTCCGGTCTTCTTGTGATAGCCATCCCTGGGCTGTCCAGGACTGGAGTTGCAGCT 1944
QY 561 LeuCys 562
Db 1945 TTGTGT 1950

RESULT 2
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
AC AAH15102;
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:13129.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPT; 2001-318749/34.

XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 13129; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC polynucleotides, all of which are used in the exemplification of the
CC present invention

XX
XX SQ Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;
Alignment Scores: 1.26e-176 Length: 2772
Pred. No.:

Score: 2981.00 Matches: 561
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.73% Indels: 0
DB: 4 Gaps: 0
US-09-787-016a-3 (1-562) x AAH15102 (1-2772)
QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 301 ATGGACGACAAAGGGGACCCGAGCAATGAGGAGGACCTTAAGGCCATCAACCCACGAGC 360
QY 21 LysGluPheArgLysThrTrpGlyPheArgGlyThrIleAlaLysArgGlyVala 40
Db 361 AAAGAGTTCCAGAAACATGGGGTTTTGGAAGACCACTATCCGCCAGGAGGGGGCGCA 420
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 421 GGGGACGCGAGGCTGACCCACTGGAGCGCCGCCACACAGCAGCAGCTGGGCTGTCC 480
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheIleuThrIle 80
Db 481 CTGGCGCGCAGTGGGAGGCGAGCCCAAGCGCACTGAGCGGCTGGAGCAGTTCCTGACCAT 540
QY 81 AlaArgArgGlyArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 541 GCGCGCGCGCGCGCGAGGAGGAGCTCTGCTCTCTCTGGAGGATTCCTGAGAGCCACG 600
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 601 TCCTGCCGCCGCCACAGACGCCGAGAGCGCTCCGAGGCGAGCGGTGGAAAGCGCTTCTGAG 660
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 661 ACCAGAGGCGCGCGCGCGAGCTCTGCTCTCAACAGCTGTGAGGAGACACAGCGCTTCTGAA 720
QY 141 LysValLysGlyArgAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 721 AAGCTGAAGAGGCGCATGACCACTGACCACTCCGATAGTCAGCAGCGATGCGCTGAC 780
QY 161 LeuLysGluLeuGlnPheArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgPro 180
Db 781 TTGAAGAGGCTTCAGAAATCGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGlyArgGlyArgGlyArgGlyArgGly 200
Db 841 CTGAAGGAGATCCAGAGTCTGCTCGGAGAGAGCGCGGAGGAGGAGGAGGAGGAGGAGT 900
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 901 GTGGGCTCCGAGGCGCGAGTGCACCTGTGGAGGGGCTCTGCGCCAGTAAAGCAGGAGCCG 960
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGly 240
Db 961 AACGATCAGGGGGTGTGTCCTGAGGCTGGGAGAGATGACAGAGAGAGTAAAGTTGGAG 1020
QY 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db 1021 AAGCGGCTCAGGACATCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 261 CysGluGlyThrAspProAsnAlaLeuTyrcysLysCysArgGlnProHisAsnAsnArg 280
Db 1081 TGTGAGGTTTACGACCCCAAGCCCTGTATGTATGATTTGCGCCAGCTTCAACACAGG 1140
QY 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
Db 1141 TTTATGATTGCTGTGACCGCTGTGAAGATGTTTTCATGGCGGATTTGTGGGCAATTCT 1200
QY 301 GluAlaArgGlyArgLeuGluArgAsnGlyValAspThrIleCysProAsnCysThr 320
Db 1201 GAGGCTCGAGGAGGCGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAACTGCACC 1260
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340

Db	1261	ATTTCGCAAGTCAGCATGAGACTCATTTCCAGAAACGGCAGATCAGCGGAAGCTAAATGG	1320
Qy	341	ATGProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnIlyssSer	360
Db	1321	AGACCTGGAGATGCTGATGGCACCAGTTGTACAAGTATAGGACCAATAGAGCAGAAGTCT	1380
Qy	361	SerGluAspGlnGlyIleLysGlyArgIleGlyLysAlaAlaAspProSerGlyLysLys	380
Db	1381	AGCGAAGACCAAGGGATAAAGGGTAGAATTTGAGAAAGCTGCAATCCAAAGTCGCAAGAG	1440
Qy	381	LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro	400
Db	1441	AAACTCAGATCTTCAGCCTGTGATAGAGGCCCTGGTGCCCTCAAAATGATATGGCCCC	1500
Qy	401	GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys	420
Db	1501	GGGTGCTGTCACGTGGCGCAGCCGACCTCGGTGTACTGCAGTAATGACTGTATCTCTCAA	1560
Qy	421	HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys	440
Db	1561	CACGCCGCGACGCAATGGAAGTTTCTAAGCTCAGGTAAAGAACAGAAAGCCCTAAA	1620
Qy	441	GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly	460
Db	1621	GAAGAAGTGAAGATGAAGCCAGAGAGCCAGCTCTTCGGAATGCGGTGCTCAGGCAGGT	1680
Qy	461	IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys	480
Db	1681	ATTAAATCTCTTCTGTGTCACAAAGAGACCCAGCTCCAGAAAAAAGAGACCAAGTGAAG	1740
Qy	481	LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer	500
Db	1741	AAGCGAGTGTGTGTCTCTGCGGAGTGAAGCACTCGGGAAGGACGAGCTTGTGAGACC	1800
Qy	501	SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla	520
Db	1801	AGCACGCGGTGCTGGGCGAGCGATCAATTAACAATGCAGTAAAGCCAGAAAAGACTGCT	1860
Qy	521	AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp	540
Db	1861	GCTCCCTCGCGGTCACTGTGTATTAATGATATGATCACTAGGGTTGGCTCTCTGGAC	1920
Qy	541	ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla	560
Db	1921	CCCTCCGGTCTTTCTGGATAGCCATCCCTGGGCGCTGTCCAGACTGGGAGTTGCAGCT	1980
Qy	561	LeuCys 562	
Db	1981	TTGTGT 1986	
RESULT 3			
ABV28029			
ID	ABV28029 standard; cDNA; 2085 BP.		
XX	AC ABV28029;		
XX	AC ABV28029;		
XX	16-SEP-2002 (first entry)		
XX	Human prostate expression marker cDNA 28020.		
XX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;		
XX	pharmacogenomic marker; gene; ss.		
OS	Homo sapiens.		
XX	WO200160860-A2.		
PN	XX		
XX	23-AUG-2001.		
PD	XX		
XX	20-FEB-2001; 2001WO-US005171.		
PF	XX		
XX	17-FEB-2000; 2000US-0183319P.		
XX	PR		

690 AAGGTGAAGAGGAGGATGACCAAGATGACACCTCGATAGTGACAGGATGGCTG-ACC 748
161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGluGluProThrGluArgPro 180
749 TTGAAGAGCTTCAGAAATCGCTT-CCACAGAGCGGGAACAGAGCCCACTCAGAGGCC 807
181 LeuLysGlyIleGlnSerArgLeuArgLysArgArgGluGluGluGluProAlaGluThr 200
808 CTGAAGGGATCAAGATCGCTCGGAGAGAGGCGCGGAGGAGGGTCCCGCCGAGACT 867
201 ValGlySerGluAlaSerAspThrValGluGluValLeuProSerLysGlnGluProGlu 220
868 GTGGCTCCAGAGCCAGTGCACACTGTGAGGGGCTCTGCCCACTAAGCAGGAGCCGAG 927
221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGlnSerLysValLeuGluGly 240
928 AACGATCAGGGGGTGTGTCCCGAGCTGGGAAGATGACAGAGAGTAAGTTGGAGGA 987
241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
988 AAGGCGGCTCAGGACATCAAGATGAGGAGCTTGAGACTTGGGCCGACCGAAGCCTGAA 1047
261 CysGluGlyThrAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArg 280
1048 TGTGAGGTTACACCCCAACCGCCCTGTATTGCTATTTGCGCGAGCTCACACACAGG 1107
281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
1108 TTTATGATTTGCTGTACCCCTGTGAGATGATGTTTCATGGCGATTTGTGGCAATTTCT 1167
301 GluAlaArgGlyArgLeuGluArgGlnGlyAspGlyIleCysProAsnCysThr 320
1168 GAGGCTCGAGGAGGCTTTTGAAGAGAAATGGGGAAGACTATATCTGCCAAACTGCACC 1227
321 IleLeuGlnValClnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLysTrp 340
1228 ATTCTCGAGTCAGGATGAGATCAATTCAGAACCGGAGATCAGCAGAGCTAATAGG 1287
341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
1288 AGACCTGGAGATCTGATGGCACCGGATTGTACAGTATAGGAACAATAGACGAGAAGTCT 1347
361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
1348 AGCGAAGACCAAGGATAAAGGGTAGAATTTGAAGAGCTCAATTCAGATGGCAAGAG 1407
381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
1408 AAATCAGATCTTCAGCTGTGATAGAGCGCTGTGCTCCCTCAAAATGATTGGCCCC 1467
401 GlyCysCysHisValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLys 420
1468 GGGTGTCTGTCAGCGGCGAGCGGACTCGGTGTACTGCAAGTAAGACTGTATCTCCAAA 1527
421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
1528 CAGCCCGCAGCAGCAATGAAGTTTCTAAGCTCAGTAAGACAGAGGCCAAAGCCTAAA 1587
441 GluLysMetLysMetLysProGluLysProSerLysProLysCysGlyAlaGlnAlaGly 460
1588 GAAAGATGAAGATGAAGCCAGAGAGCCAGCTTCCGAATGCGGTCTCAGGCGAGGT 1647
461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
1648 ATTAATAATCTTCTGTGACACAGAGACCTCAGTCCAGAAAAAAGAGACCAAGTGAAG 1707
481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
1708 AAGCAGTGGTGTCTCTCGCGGAGTGAAGCACTCGGGAGGAGGAGCAGCTTGTGAGAGC 1767
501 SerThrProSerTrpAlaSerAspHisAsnTyrrAsnAlaValLysProGluLysThrAla 520
1768 AGCAGCGGCTGTGGGCGAGGATCACAAATTCATGCTAAGCTAAGCCAGAAAGACTGCT 1827

QY 521 AlaProSerProSerLeuLeuTyrlsCysMetTyrlHisLeuGlyValGlyLeuLeuAsp 540
Db 1828 GCTCCCTCGCGTCACTGTCTATATAATGATGATACCTAGGGTTGGCTCTCTGCAC 1887
QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
Db 1888 CCTCCCGCTTCTTCTGATAGC-ATCCCTGGGCTCTCCAGGACTGGAGTTGAGCT 1946

RESULT 4
ABV27935
ID ABV27935 standard; cDNA; 2085 BP.
XX
AC ABV27935;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 27926.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0213007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5757-5758; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 7,23e-162 Length: 2085
Score: 2742.00 Matches: 551
Percent Similarity: 98.39% Conservative: 0
Best Local Similarity: 98.39% Mismatches: 9
Query Match: 91.74% Indels: 9
DB: 5 Gaps: 0

US-09-787-016A-3 (1-562) x ABV27935 (1-2085)

1 MetAspLysGlyValSerProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 276 ATGACGCAAAAGCGACCCCGCAATGAGGGGCACTTAAGGCCATCAAAACCCACCCAGC 335
 21 LysGluPheArgLysThrTyrGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 336 AAAGAGTTTCAGGAAACATGGGTTTTCGAAGGACCACTATCGCCAAAGCGAGAGGGCGCA 395
 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 396 GGGGACGGAGGCTGACCCACTGAGCGCGCA-CCCCCAGACGACGAGCTGGGCTGTCC 454
 61 LeuArgSerGlyValGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 455 CTGCGGCGCAGTGGGAGGAGCGGCCAAGCGCACTGAGCGCTGGAGCACTTCCTGACCAATT 514
 81 AlaArgArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 515 GCGCGCGCGCGCGGAGGAGGAGCTGCTGCTCCCTGGAGGATTCCTGGAGGCCACG 574
 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 575 TCTCTCCCGCG-ACAGACGCCGAGACAGC-TCCGAGGSCAGCGTGGAAAGCGCTTCTGAG 632
 121 ThrArgSerGlyProGlnSerThrAlaValLysGluArgProAlaSerSerGlu 140
 633 ACCAGAAGCGGCGCCGAG-TCTGCTTCCACAGT-GTGAAGGACAGCACCAGC-TCCTTCTGAA 689
 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspGlyLeuThr 160
 690 AAGGTGAAGAGGGGAGTACACCGATGACACTCCGATAGTGACAGCGATGGCTG-ACC 748
 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
 749 TTGAAGAAGCTTCAGAAATCGCTT-CGCAAGAAAGCGGGAACAGAGCCCACTGAGAGGCC 807
 181 LeuLysGlyIleGlnSerArgLeuArgLysArgArgGluGluGluGlyProAlaGluThr 200
 808 CTGAAGAAGGATCCAGAGTCCCTGCGAAGAGAGCGCGGAGAGGGTCCCGCCGAGACT 867
 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerIleGlnProGlu 220
 868 GTGGGCTCCGAGGCCAGTCACACTGTGGAGGGCGCTCTGCCCAAGTAAGCAGAGGCCGAG 927
 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 928 AACGATCAGGGGGTGTGTCCAGGCTGGGAAGATGACAGAGAGTAAGTTGGAGGA 987
 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 988 AAGGGCGCTCAGACATCAAGATGAGGAGCCTGGAGACTTGGCGCGACCGAAGCCTGAA 1047
 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 1048 TGTGAGGGTTAGACCCCAAGCCCTGTATTGTATTTGGCGCAGCTCACACACACAG 1107
 281 PheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300
 1108 TTATGATTTGTGTGACCGCTGTGAAGATGCTTTCATGGCGATTTGTGGGCAATTTCT 1167
 301 GluAlaArgGlyArgLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 1168 GAGGCTCGAGGGAGGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1227
 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
 1228 ATTCTGCAGTGCAGATGAGACTCAATCAGAAACGGCAGATCAGCAGGAGTAATGG 1287
 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 1288 AGACCTGGAGATCTCTATGGCACCGGATTTGTACAAGTATAGGACAAATAGACGAGAGTCT 1347

361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 1348 AGCGAAGACCAAGGGATAAAAGGCTAGAAATTGAGAAAGCTCAATTCCACTGGCAGAGAG 1407
 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 1408 AAATCTCAGATCTTCCAGCTGTGATAGAGCGCTGGTGCCTCAAAATGTATTGGCCCC 1467
 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 1468 GGGTGTCTGTCACTGGCGCAGCCCGACTCGGTGTACTGCAGTAACTGTATCTCTCAAA 1527
 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 1528 CACCCCGCAGCAGCATGAAGTTCTTAAGCTCAGGTAAGAACACAGAGCCAAAGCCTAAA 1587
 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 1588 GAAAGATGATGAATGAAGCCAGAGAGCCAGTCTTCCGAAATGCGGTGTCTCAGGCAGGT 1647
 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 1648 ATTAATAATCTCTTCTGTGCAAGAGACAGCAGCTCCAGAAAAAAGAGACACACAGTGAAG 1707
 481 LysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 1708 AAGCCAGTGGTGGTCCCTCTCGCGAGTGAAGCACTCCGGAAGGAGCAGCTTGTGAGAGC 1767
 501 SerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 1768 AGCAGCGCGTGTGGGCGAGCATCAATTAATGCACTGAAGCCAGAAAAAGAGACTGCT 1827
 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 1828 GCTCCCTCGCGCTCACTGTTGTATTAATGTATGTATCACTAGGGGTGGCTCTCTGGAC 1887
 541 ProSerArgSerPheThrIleAlaIleProThrAlaCysProGlyLeuGlyValAlaAla 560
 1888 CCCTCCCGTCTTTCTGTGATAGC-ATCCCTGGGCGCTGTCCAGGACTGGAGTTGCAGCT 1946

RESULT 5

ABV22095 standard; cDNA; 2085 BP.

 XX ABV22095;
 AC ABV22095;

 DT 13-SEP-2002 (first entry)
 DE Human prostate expression marker cDNA 22086.

 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.

 PN WO200160860-A2.
 PD 23-AUG-2001.

 PF 20-FEB-2001; 2001WO-US005171.
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 3785; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Alignment Scores:
 Pred. No.: 7,23e-162 Length: 2085
 Score: 2742.00 Matches: 551
 Percent Similarity: 98.39% Conservative: 0
 Best Local Similarity: 98.39% Mismatches: 9
 Query Match: 91.74% Indels: 9
 DB: 5 Gaps: 0

US-09-787-016A-3 (1-562) x ABV22095 (1-2085)

QY 1 MetAspIysGlyValProSerAsnGluAlaProIysAlaIleLysProThrSer 20
 DB 276 ATGACACAAAGCGCCGACATGAGAGGACCTAAGGCCATCAAAACCCACGCG 335
 QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 DB 336 AAGAGTTTCAGGAAACATGGGTGTTTCGAAGGACCACTATGCGCAAGCGAGGGGCGCA 395
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnLeuGlyLeuSer 60
 DB 396 GGGGACCGGAGGCTGACCCACTCGAGCGCGCA-CCCCACAGCAGCAGCTGGCGCTGTC 454
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 455 CTGCGGCGCAGTGGAGGCGACCCACGCGCACTGAGCGCGTGGAGCAGTTCTCTGACCAT 514
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 515 GCGCGGCGCGCGCGAGGAGCATGCTGCTCTCCCTGGAGGATTCCTGTGAGCCACG 574
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 575 TCTTGGCCCCGC-ACAGACGCGGAGCAGC-TCGAGGCGCAGCGTGGAAAGCGCTTCTGAG 632
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 DB 633 ACCAGAAGCGCCCCAG-TCGTGCTTCCACAGT-GTGAAGGAGAACCCAGC-TCCTTCTGAA 689
 QY 141 LysValLysGlyArgAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 690 AAGGTGAAGAGGGGATGACCAAGATGACACCTCCGATAGTACAGCATGGCTG-ACC 748
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
 DB 749 TTGAAGAGAGCTTCAGATCGCTT-CGCAAGAGCGGGAAACAGGAGCCCACTGAGAGGCC 807
 QY 181 LeuLysGlyLeuGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
 DB 808 CTGAAGAGGAGATCCAGATCGCTTCCGAGAGAGCGCGCGGAGGAGGTCCCGCCGAGACT 867
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220

RESULT 6
 ABV22192

DB 868 GTGGGCTCCGAGCCAGTCACTGTGGAGGGCGTCTCCGCCAGTAAAGAGGAGCCGAG 927
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 DB 928 AAGGATCAGGGGTGCTGCTCCAGCGCTGGGAAAGATGACAGAGAGAGTAAGTTGGAGGGA 987
 QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 DB 988 AAGGCGGCTCAGGACATCAAGATGAGGAGCTTGGAGACTTGGGCCGACCGAAGCTGAA 1047
 QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 DB 1048 TGTGAGGGTTACGACCCCAACCGCTGTATGCTATTCGATTCGCGCCAGCTCACACAACAG 1107
 QY 281 PheMetIleCysCysAspArgCysGluGlnTrpPheHisGlyAspCysValGlyIleSer 300
 DB 1108 TTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTCATGGCGATTTGTGGCGCATTTCT 1167
 QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 DB 1168 GAGGCTCGAGGGAGGGCTTTTGGAAAGGATGGGAAGACTATATCTGCCCAATGCAACC 1227
 QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
 DB 1228 ATTCTCAAGTGCAGGATGAGACTCATTCAGAAACGCGCAGATCAGCAGGAAGCTAAATGG 1287
 QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 DB 1288 AGACCTGGAGATGCTGATGCGACCGATTGTACAGTATAGGAACAATAGACAGAGAGTCT 1347
 QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 DB 1348 AGCGAAGACCAAGGATAAAGGGTAGAATTGAGAAAGCTGCAATCCAGTGGCAAGAG 1407
 QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 DB 1408 AAATCAAGATCTTCAGGCTGTGATAGAGCGCCCTGGTGGCTTCAAAATGATTTGGCCCC 1467
 QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 DB 1468 GGGTGTGTGTCAGTGGCGAGCCCGACTCGGTGTACTGTCAGTAATGACTGTATCTCTCAA 1527
 QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 DB 1528 CACGCGCAGCGCAATGAAGTTTCTAAGCTCAGGTAAAGAAACAGAGAGCCCTTAA 1587
 QY 441 GluLysMetIleMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 DB 1588 GAAGAATGAGATGAGCAGCAGAGAGCCCACTCTTCCGAATGCGGTGCTCAGGAGGT 1647
 QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 DB 1648 ATTAAATCTTCTGTGTCACAGACAGCAGCTCCAGAAAAAAGAGAGACACAGTGAAG 1707
 QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysCysGluSer 500
 DB 1708 AAGCGAGTGTGTCTCCGCGGAGTGAAGACTCGGAGAGAGAGCAGCTTGTGAGAGC 1767
 QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 DB 1768 AGCAGCGCTGCGTGGCGAGCGATCAATATCAATGTCAGTAAAGCCAGAAAAAGACTGCT 1827
 QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 DB 1828 GCTCCCTGCGCGTCACTGTGTATAATGATGATATCACTAGGGGTGGCCCTCTGGAC 1887
 QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 DB 1888 CCTCTCCGTTCTTCTGCGATAGC-ATCCCTCTGGCGCTGTCAGGACTTGGAGTTCAGCT 1946

ID ABV22192 standard; cDNA; 2085 BP.
XX AC ABV22192;
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 22183.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX OS pharmacogenomic marker; gene; ss.
XX XX Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US0051171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PA Schlegel R, Endege WO, Monahan JB;
XX PI WPI; 2001-662795/76.
XX DR Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.
XX BS Claim 1; Page 3816; 11750pp; English.
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX SQ Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 1.48e-161 Length: 2085
Score: 2737.00 Matches: 550
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 10
Query Match: 91.57% Indels: 9
DB: 5 Gaps: 0

US-09-787-016a-3 (1-562) x ABV22192 (1-2085)

QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
DB 276 ATGAGACACAAAGCGCCGAGCAGTATGAGGAGGACCTATAGCCCATCAACCCACGAC 335
QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
DB 336 AAAGAGTTTCAGGAAACATGGGGTTTTCGAAGACCACTATGCCAAGGAGGAGGCGCA 395
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnLeuGlyLeuSer 60

396 GGGAGCGCGAGGCTGACCCACTGGAGCCGCA-CCCCACAGCAGCAGCTGGGGCTCTCC 454
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 455 CTGCGCGCAGTGGAGGCGAGCCACAGCGACTGAGGCGCTGGAGCAGTTCTTGACCAT 514
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 515 GCGCGGCGCGCGCAGGAGGAGCATGCTGTCTCTCCCTGGAGGATTCTGGTGGAGCCACG 574
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 575 TCCTGCCCCCG-ACAGACCGCGAGCAGC-TCGAGGGCGAGCTGGAAAGCGCTTCTGAG 632
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 633 ACCAGAGCGGCCCCAG-TCTGCTTCCACAGT-GTGAAGGAACCCAGC-TCTTCTGAA 689
QY 141 LysValLysGlyLysAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 690 AAGGTGAAGAGGAGGAGTACCCAGATGACACCTCCGATAGTGACAGGATGGTGG-ACC 748
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
DB 749 TTGAAGAGCTTCAGANTGCTT-CGCGAAGCGGGAACAGGAGCCCACTGAGAGGCC 807
QY 181 LeuLysGlyLysGlnSerArgLeuArgLysArgGluGluGluGluGluProAlaGluThr 200
DB 808 CTGAAGCGGATCCAGAGTCCGCTCGGAAGAGCGCGGAGGAGGGTCCCGCGAGCT 867
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
DB 868 GTGGCTCCGAGGCGCAGTGACACTGTGGAGGCGCTCTGCCAGTAAGCAGAGCCGAG 927
QY 221 AsnAspGlnGlyValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
DB 928 AACGATCAGGGGGTGTGTCCAGGCTGGGAAGATGACAGAGAGTAAGTTGAGGAGGA 987
QY 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
DB 988 AAGCGGCTCAGGACATCAAGATGAGGAGCTGGAGACTGGGCCGACCGAGCTGAA 1047
QY 261 CysGluGlyTrpAspProAsnAlaLeuTyrcysLysCysArgGlnProHisAsnAsnArg 280
DB 1048 TGTGAGGGTTACGACCCCAACGCGCTGTATTGCAATTCGCCCGAGCCCTCACAACACAGG 1107
QY 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
DB 1108 TTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTCATGGCGATTGTGTGGCAATTTCT 1167
QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCysThr 320
DB 1168 GAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1227
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
DB 1228 ATTTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGACTAAATGG 1287
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
DB 1288 AGACCTCGAGATGCTGTATGGCACCGATTGTACAGTAAGCAACATAGAGCAGAGTCT 1347
QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
DB 1348 AGCGAAGCAAGGAGTAAAGGTAGATTCAGAAAGCTGCATAATCCAGTGGCAGAGAG 1407
QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
DB 1408 AAACCTCAGATCTTCACGCTGTGTAGAGGCGCTGTGTGCTCAAAATGATTATGGCCCC 1467
QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLys 420
DB 1468 GGCTGCTGTGCTGCGCGAGCCCGACTCGGTGTACTGCGTAATGACTGTATCTCAAA 1527

Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1528 CAGCGCCACCGACCAATGAAGTTCTTAAGCTCAGGTAAGACACAGAGCCAAAGCCTAAA 1587
 Qy 441 GlnLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1588 GAAAGATGAAGATGAAGACGAGAGCCCAAGTCTTCGAAATGCGGTCTCAGCGAGGT 1647
 Qy 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1648 ATTAATATCTCTCTGTGCAACAGACCCAGCTCCAGAAAAAAGAGACACACAGTGAAG 1707
 Qy 481 LysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysCysGluSer 500
 Db 1708 AAGSCAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGGAAGGAGCAGCTTTGGAGAGC 1767
 Qy 501 SerThrProSerTrpAlaSerAspHisAsnTrpAsnAlaValLysProGluLysThrAla 520
 Db 1768 AGCAGCGCGCTGTGGCGAGCGATCACATTTACATGCAATGATTAAGCCAGAAAGACTGCT 1827
 Qy 521 AlaProSerProSerLeuLeuTyrlLysCysMetTrpHisLeuGlyValGlyLeuLeuAsp 540
 Db 1828 GCTCCCTCGCGCTCACTGTTGTATATAATGTATGTATCATCCTAGGGGTGGCCCTCTGGAC 1887
 Qy 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 Db 1888 CCCGCCCGCTCTCTGGATAGC-ATCCCTGGGCGCTGCCAGCTGGGAGTTGCAGCT 1946
 RESULT 7
 AAK52368
 ID AAK52368 standard; cDNA; 7838 BP.
 AC AAK52368;
 XX
 XX 06-NOV-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 913.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200157190-A2.
 EN
 XX 09-AUG-2001.
 FD
 XX 05-FEB-2001; 2001WO-US004098.
 PF
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00520325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 XX WPI; 2001-476283/51.
 DR P-FSDB; AAM79235.
 DX
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX

PS Claim 1; Page 3007-3013; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.6e-143 Length: 7838
 Score: 2449.00 Matches: 477
 Percent Similarity: 90.75% Conservative: 4
 Best Local Similarity: 90.00% Mismatches: 8
 Query Match: 81.93% Indels: 41
 DB: Gaps: 4
 US-09-787-016A-3 (1-562) x AAK52368 (1-7838)
 Qy 37 ArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGln 56
 Db 528 CGCAGAGCCAGGAGAGGCT---GGGAGCG---CAGCGGCCAGACCGCGAGCGCG 581
 Qy 57 LeuGlyLeu-SerLeuArgSerGlyArgGlnProLysArgThrGluArgValGluG 76
 Db 582 GGAGGCGAGGTTCCGACGT-----GGGAGCGAGCCCAAGCGCACTGAGCGCGTGGAGCA 635
 Qy 76 nPheLeuThrIleAlaArgArgGlyArgArgSerMetProValSerLeuGluAspSe 96
 Db 636 GTTCCTGACCATTTGCGCGCGCGCGCGAGAGAGCATGCTGTCTCTCCCTGGAGGATTC 695
 Qy 96 rGlyGluProThrSerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValG 116
 Db 696 TGGTGAGCCCACTGCTCTGCGCGCGCGCGAGAGAGCATGCTGTCTCTCCCTGGAGGATTC 755
 Qy 116 uSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgPr 136
 Db 756 AAGCGCTTCTGAGACCAAGAGCGCGCGCGCGAGAGAGCATGCTGTCTCTCCCTGGAGGATTC 815
 Qy 136 oAlaSerSerGluLysValLysGlyLysAspHisAspHisAspThrSerAspSerAspSe 156
 Db 816 AGCTCTCTCTGAAAAGGTGAAGAGGGGATGACACGATGACACCTCCGATAGTACAG 875
 Qy 156 rAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluPr 176
 Db 876 CGATGGCTGACCTTGAAGAGCTTCAGAAATCGCTTTCCGAGGAGAGCGGAAACAGAGGCC 935
 Qy 176 oThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysArgGluGluG 196
 Db 936 CACTGAGAGGCCCTCTGAAGAGGATCCAGATGCTGCTGCGAGAGAGCGCGCGAGGAGGG 995
 Qy 196 yProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLy 216
 Db 996 TCCCGCGAGACTGTGGGCTCCGAGGCGAGTGCACACTGTGGAGGGCGCTCTCCCTCCAGTAA 1055
 Qy 216 sGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspArgGluSe 236
 Db 1056 GCAGGAGCCCGAGAACCATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGAG 1115
 Qy 236 rLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyAr 256
 Db 1116 TAGTTGAGGGGAAAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGACTTGGGCGCG 1175
 Qy 256 gProLysProGluCysGluGlyTrpAspProAsnAlaLeuTyrlCysIleCysArgGlnPr 276

Db 1176 ACCGAAGCCTGAATGTAGGGTTAGACCCCAACCCCTGATTTCATTTGCGCCGACCC 1235
Qy 276 cHisenAenArgPheMetIleCysCysAspArgCysGluGluTtpPheHisGlyAspCy 296
Db 1236 TCACACACACAGTTTATGTTGCTGTGACCCCTGTGAAGATGGTTTCATGCGCATG 1295
Qy 296 sValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCy 316
Db 1296 TGTGGGCAATTCAGGCTCGAGGAGGCTTTTGAAAGGAATGGGGAAGACTATATCTG 1355
Qy 316 sProAenCysThrIleLeuGluValGluAspGluThrHisSerGluThrAlaAspGlnG 336
Db 1356 CCCAACTGCACCATTCCTCAAGTCAGGATGAGACTCATTCAGAAACGGCAGATGACGA 1415
Qy 336 nGluAlaIysTtpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrI 356
Db 1416 GGAAGCTAATGGAGACCTGGAGATGCTGATGGCACCAGTGTACAGATATAGGAACAAT 1475
Qy 356 eGluGlnIysSerSerGluAspGlnGlyIleIysGlyArgIleGluIysAlaAlaAsnPr 376
Db 1476 AGAGCAGAAAGTCTAGCGAAGACCAAGGGATAAGGGTAGAATGAGAAAGCTGCAATCC 1535
Qy 376 oSerGlyLysLysLysLeuIysIlePheGlnPro----- 387
Db 1535 AAGTGGCAAGAGAAACTCAGATCTTCCAGCTGTGTCCCGGCTGTCCCCACCCAGCT 1595
Qy 387 ----- 387
Db 1596 GCCTGTCTCTGGCAGGTATTGAAATTCGTGTCTAGAGCATCTTGCCTTCACACT 1655
Qy 388 -----ValIleGluAlaProGlyAlaSerLysCysIleGlyPr 400
Db 1656 CTGTGCACTGATAGCTGCAGGTGATAGAGGCGCTGTGTCTCAATGATATTGGCCC 1715
Qy 400 oGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLy 420
Db 1716 CGGGTGTCTGCTGAGTGGCGCAGCCGACTCGGTGTACTGCAGTAATGACTGTATCTCAA 1775
Qy 420 sHisAlaAlaThrMetIysPheLeuSerSerGlyLysGluGlnIysProIysProLy 440
Db 1776 ACAGCGCGCAGCAGCATGAATGAATTTCTAGCTCAGTTAAGACACAGACCAAGCTTAA 1835
Qy 440 sGluLysMetLysMetIysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGl 460
Db 1836 AGAAAGATGAAGATGAAGCAGAGAGCCAGCTTCCGAAATGCGGTCTCAGGACAG 1895
Qy 460 yIleLysIleSerSerValHisIysArgProAlaProGluLysLysGluThrThrValLy 480
Db 1896 TATTAAATCTCTTCTGTGCACAGACAGACCGCTCCAGAAAAAAGAGACCAAGTGAA 1955
Qy 480 sLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSe 500
Db 1956 GAAGGAGTGTGTGCTTCTGCGGAGTGAAGCAGCTCGGAGAGGAGCAGCTTGTGAGAG 2015
Qy 500 rSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValIysProIulysThrAl 520
Db 2016 CAGCAGCCGCTGTGTGGCGAGCGATCAATTAATGATGATGAGCAAGCAAGAAAGACTGC 2075
Qy 520 aAlaProSerProSerLeuLeuTyrLys 529
Db 2076 TGCTCCCTCGCCGTCACCTGTGTATATAA 2103
RESULT 8
ID AAK53352
XX AAK53352 standard; cDNA; 7290 BP.
XX AC AAK53352;
XX XX
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 2881.
XX

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation; ss.
Homo sapiens.
PW WO200157190-A2.
PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US004098.
XX 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSB-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR P-PSDB; AAM80219.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.
XX Claim 1; Page 5053-5055; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52592) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication;
XX Sequence 7290 BP; 1762 A; 1953 C; 1975 G; 1600 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,88e-142 Length: 7290
Score: 2440.00 Matches: 475
Percent Similarity: 91.06% Conservative: 4
Best Local Similarity: 90.30% Mismatches: 6
Query Match: 81.63% Indels: 41
DB: Gaps: 4
US-09-787-016A-3 (1-562) x AAK53352 (1-7290)
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeu-Se 60
Db 5 GGAGAGGCT---GCGGAGCGG---CAGCGCGCCAGACCGCGGAGCGCGGAGCGAGTTC 58
Qy 60 rLeuArgArgSerGlyArgGlnProIysArgThrGluArgValGluGlnPheLeuThrI 80
Db 59 CGCACGT-----GGGAGGCGAGCCCAAGCGCATGTAGCGCGTGGAGCATTTCTTGACCAT 112
Qy 80 eAlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProTh 100
Db 113 TGGCGCGCGCGCGCGCGCGCGAGGAGCATGCTGTCTCTCTCTGGAGGATTCCTGCTGAGCCAC 172

QY 100 rSerCysProAlaThrAspAlaSerGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 173 GTCTGCCCCGACAGACGCGGAGCAGCTCCGAGGGCGCGTGGGAAGGCTTCTCA 232
 QY 120 uThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 233 GACCAGAGCGGCCCCAGTCTGCTTCCACAGCTGTGAAGGACGACCACTCTTCTCA 292
 QY 140 uLysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuTh 160
 Db 293 AAAGGTGAAGGAGGAGGATGACACCATGACCTCCGATAGTGAAGGATGGCCCTGAC 352
 QY 160 rLeuLysGluLeuGlnAsnArgLeuArgGlyAspArgGluGlnGluProThrGluArgPr 180
 Db 353 CTTGAAGAGCTTCAGATCGCTTCCGAGGAGCGGACAGGAGGCCCACTGAGAGGCC 412
 QY 180 oLeuLysGlyLeuGlnSerArgLeuArgLysLysArgArgGluGluGluGluProAlaGluTh 200
 Db 413 CTTGAAGGATCCAGAGTCCCTGCGGAGAGAGCGCGGAGGAGGCTCCGCGCGAGAC 472
 QY 200 rValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 473 TGTGGGCTCCGAGGCCAGTGAACCTGTGGAGGGCGTCTGCCCCAGTGAAGGAGGCCGGA 532
 QY 220 uAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 240
 Db 533 GAACGATCAGGGGCTGTGTCCAGGCTGGGAAAGATGACAGAGAGAGTAAAGTTGGAGGG 592
 QY 240 yLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 Db 593 AAAGGGGGCTTCAGACATCAAGATAGGAGGCTGGAGACTTGGGGCCGACCGAGGCTGA 652
 QY 260 uCysGluGlyLysAspProAsnAlaLeuTyrcysIleCysArgGlnProHisAsnAsnAr 280
 Db 653 ATGTGAGGTTACGACCCCAAGCGCTGTATTGCTATTTGCGCGCAGGCTCACAAACACAG 712
 QY 280 gPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSe 300
 Db 713 GTTTATGATTGCTGTGACCGCTGTGAAGATGGTTTCATGGCGATGTGTGGCGCATTC 772
 QY 300 rGluAlaArgGlyValGluLeuGluArgAsnGlyGluAspTyIleCysProAsnCysTh 320
 Db 773 TGAGGCTCGAGGGAGGCTTTTGGAAAGGATGGGAGAGCTATATCTGCCCAACTGCAC 832
 QY 320 rIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTr 340
 Db 833 CATCTGCAAGTGCAGGATGAGATCTCATTCAGAAACGGCAGATCAGCAGGAGGCTAAATG 892
 QY 340 pArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSe 360
 Db 893 GAGACCTGGAGATGCTGATGGCCGATTTGTACAGTATAGGACATATAGCAGCAGAGTTC 952
 QY 360 rSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLy 380
 Db 953 TAGCGAAGACCAAGGGATAAGGGTAGAATTAGAAAGCTGCAATCCAAGTGCAGAA 1012
 QY 380 sLysLeuLysIlePheGlnPro----- 387
 Db 1013 GAAACTCAGATCTTCAGCGCTGTGTCGGGCGCTGTCCCGACCGAGCTGCTGTCTCTG 1072
 QY 387 ----- 387
 Db 1073 GCAGGTATTGGAATTCGTGTCTAGAGCATCTCTCGCTTCACACTCTTGCACTGCAT 1132
 QY 388 -----ValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHi 404
 Db 1133 TAGCTGCAAGGTGATGAGGCGCTGTGTGCTCAAAATGTATTGGCCCCGGGTGCTGTC 1192
 QY 404 sValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLysHisAlaAla 424
 Db 1193 CTTGGGCGCAGCCGACTCGGTGACTGACTGATGATGATCTCTCAACACGCGCGAGC 1252

QY 424 aThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysGluLysMetLy 444
 Db 1253 GACATGAAGTTCTTAAGCTCAGGTAAAGAACAGAGCCAAAGCCTAAAGAAAGATGAA 1312
 QY 444 sMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSe 464
 Db 1313 GATGAAGCCAGAGAACCCAGTCTTCCAAATGCGGTGCTCAGGCAGGTATTAAATCTC 1372
 QY 464 rSerValHisLysArgProAlaProGluLysLysGluThrThrValLysLysAlaValVa 484
 Db 1373 TTCTGTGCACAGAGACCCAGCTCCAGAAAAAAGAGAGACCCAGTGAAGAGGCGAGTGGT 1432
 QY 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSe 504
 Db 1433 CGTCCCTGCGCGAGTGAAGCACTCGGAGAGGAGCGCTTGTGAGAGCAGACGCCGCTC 1492
 QY 504 rTrpAlaSerAspHisAsnTyrcysAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 1493 GTGGGCGAGCGATCAACATTACATGCAATTAAGCCAGAAAGACTGCTGCTCCCTCGCC 1552
 QY 524 oSerLeuLeuTyrcys 529
 Db 1553 GTCACCTGTGTATATA 1568
 RESULT 9
 AAS86576
 ID AAS86576 standard; cDNA; 2332 BP.
 XX AC AAS86576;
 XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #22380.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 XX NC0200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG22389.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 22380; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological actions. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.73e-132 Length: 2332
 Score: 2262.00 Matches: 477
 Percent Similarity: 79.55% Conservativity: 21
 Best Local Similarity: 76.20% Mismatches: 48
 Query Match: 75.68% Indels: 81
 DB: 5 Gaps: 12

US-09-787-016A-3 (1-562) x AAS86576 (1-2332)

QY 1 MetAspHisGlyAspProSerAsnGluAlaProLysAlaLeuProThrSer 20
 DB 391 ATGGACGCAAGGCGACCCGAGCAATGAGAGCGCCTAAGGCCATCAACCCACGAC 450
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
 DB 451 AAAGAGTTTCAGGAAACATGGGGTTTCGAAAGACCACTATCCCAAGCGAGAGCGCGCA 510
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 DB 511 GGGACCGGAGGCTGACCCACCTGGAGCCGCCACCCCGACGAGCGCTGGCCCTGCTCC 570
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 571 CTGGGGCGCAGTGGAGGCGGACGCCAAGCGCACTGAGCGCGTGGAGCAGTTCTGACCAT 630
 QY 81 AlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 631 GCGCGGCGCGCGGAGAGAGAGCTGCTCTCTCTGGAGATTCTGGTGGAGCCGCG 690
 QY 101 SerCysProLalaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 691 TCTTGCCCGCCGACAGACGCGGACAGACCTCGGAGGCGAGCGTGGGAAAGCGCTTCTGAG 750
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
 DB 751 ACCAGAGCGGCGCCCGCTGCTGCTCCAGCTGTGAGGAGACGACAGCTTCTCTGAA 810
 QY 141 LysValLysGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 811 AAGGTGAAGAGGAGGAGATGACCAAGATGACCTCCGATAGTCAAGCGATGCGCTGACC 870
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 DB 871 TTGAAAGCTTTCAGATCCCTTCGCGAGAGCGGGAACAGGAGCCCACTGAGAGGCC 930
 QY 181 LeuLysGlyLeuGlnSerArgLeuArgLysArgLysArgGluGluGluProAlaGluThr 200
 DB 931 CTGAAAGGATCCAGATGCTGCTGCGGAGAGCGGCGGAGGAGGCTCCCGCGAGACT 990
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 DB 991 GTGGGCTCCGAGGCGAGTGAACCTGTGAGGGCGTCTGCGCCAGTAGGAGAGGCCGAG 1050
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGlnSerLysLeuGluGly 240
 DB 1051 AAGCATCAGGGGGTGTGTCTCCAGCTGGGAAAGATGACAGAGAGAGTAGTAGTGGAGCA 1110
 QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260

DB 1111 AAGGGCGCTCAGGACATCAAGATGAGGAGCGCTTGGAGACTTTGGCGGCGGAGCGCTGA 1170
 QY 260 uCysGluGlyTrpAspProbenAlaLeuTrpCysIleCysArgGlnProHisbenAr 280
 DB 1171 ATGTGAGGGTTACGACCCCAACGCCCTGTATTGTCAATTGGCGGCGGCTCACACACAG 1230
 QY 280 gPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyLys 300
 DB 1231 GTTATGATTGCTGTGACCGCTGTGAGAGATGGTTTCATGGCGATTGTGGCATTC 1290
 QY 300 rGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTrpIleCysProbenCys 320
 DB 1291 TGAGGCTCGAGGAGGCGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCGCAACATG 1350
 QY 320 rIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 340
 DB 1351 CATCTGCAAGTGCAGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGAGACTTAA 1410
 QY 340 rp-ArgPro-GlyAspAla-AspGlyThrAsp-CysThrSerIleGlyThrIleGluGln 358
 DB 1411 CGGAGACCTCGGAGATGCTTGTATGGCACCGGATTTTACAAAGTATAGGGAACATTAG 1470
 QY 359 Lys-----SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnPro 376
 DB 1471 AGAAGTCTTAGCCGACACCCAGGGGTTAAGGGTAGAATTGAGAAAGCTGCATATCCA 1530
 QY 377 SerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 387
 DB 1531 AGTGGCAAGAGGAACCTCAAGATCTTTCAGGCGCTGTATTCGAAATTCCTGTCTAGAAG 1590
 QY 388 -----ValIleGluAlaProGlyAl 394
 DB 1591 CATCTGCGCTTCACACTCTTGCACCTAGCTAGCTGCAAGTGTATAGAGCGGCTGGTGC 1650
 QY 394 aserLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTrpCys 414
 DB 1651 CTCAAAATGATTTGGCGCCCGGGTGTCTCACTGGCGGCGAGCCGCTGCTGCTGCTG 1710
 QY 414 rAsnAspCysIleLeuLysHisAlaAlaLarMetLysPheLeuSerSerGlyLysG 434
 DB 1711 TAATGACTATCTCTAACAACGCGCGAGCGACATGAGTCTTCTAGCTCAGTAAGA 1770
 QY 434 uGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSerLeuPro 454
 DB 1771 ACAGAGCCAAAGCTTAAGAAAGATGAAGATGAAGCCAGAGAGCCAGGCCAGTCTTCG 1830
 QY 454 sCysGlyValaGln----- 458
 DB 1831 ATGGCGTGTCTCAGCTGTCCAGTGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1890
 QY 459 -AlaGlyIleLysLysSerSerValHisLysArgProAlaProGluLysLysLysLys 478
 DB 1891 ATCAGGGTGTGAGGTG---ACTCTGTGCGAGCGCGCTTTCACGAGCGCACAG---TG 1941
 QY 478 rValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAla-- 497
 DB 1942 GTTGAGATTCTCAGTTTG-----CACGGGGAACCTGCGGCGCCACCGCGGCGCCA 1992
 QY 498 ----CysGluSerSerThrProSerTrpAlaSerAspHisAsnTrpAsnAlaValLysPr 516
 DB 1993 CCGAATACGCAAGCTCAACAAAGTACTGGGGAGCGCTGCCACTGTATGCTTTGAGGCC 2052
 QY 516 o-----GluLysThrAlaAlaProSerProSerLeuLeuTrpLysCysMetTy 532
 DB 2053 CTGCGCCCTGCTCCCTGGCTGGCCACGAGCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 2095
 QY 532 rHisLeuGlyValGlyLeuLeuAspProSerArgSerPhe----- 545
 DB 2096 -----GTAGGGG---CTGGATCCCTGCCCGCGCATGGCGTGCACCTCAACAGAGCCC 2143
 QY 546 -----TrpIleAlaIleProTrp-----AlaCysProGlyLeu---GlyValAlaAl 560
 DB 2144 TGTCCCTGGCCATCTGTCTCCCTGGCAGCGCGCTCTGTCTCTCTCTCTCTCTCTCTCT 2203

QY 560 aleu 561
Db 2204 TCTG 2207

RESULT 10
AAZ90579
ID AAZ90579 standard; cDNA; 2867 BP.
AC AAZ90579;
XX

19-JUN-2000 (first entry)
XX Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
DT
XX Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;
XX autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW antirheumatic; antiinflammatory; antiproliferative; se.
XX

OS Mus sp.
XX
XX Key Location/Qualifiers
FH 232..2076
CDS /tag= a
FT /product= "DIO-1"
FT
FT
XX
XX
PN W0200015787-A1.

XX
XX 23-MAR-2000.
XX
XX 10-SEP-1999; 99WO-GB003019.
XX
XX 10-SEP-1998; 98SE-00003069.
PR 17-SEP-1998; 98US-0100873P.
XX
XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANN/) BANKERMAN D G.
XX
XX Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
PI
XX WPI; 2000-271426/23.
DR P-PSDB; AAY67580.
XX

New DNA encoding human and murine death inducer-obliterator 1 polypeptides, useful in the treatment of cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors and hyperproliferative skin disorders.

XX Claim 3; Fig 1B; 27pp; English.

XX The invention provides nucleic acids encoding the human and murine death inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be expressed by standard recombinant methodology. The DIO-1 polypeptides, agonists and antagonists are used as a medicament for treating diseases characterized by an alteration in cell death or by hyperproliferation, e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors or hyperproliferative skin disorders. They are also useful in the treatment of metabolic, proliferative or inflammatory conditions. The present sequence represents a cDNA encoding the murine DIO-1 polypeptide

XX
XX
SQ Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,75e-122 Length: 2867
Score: 2112.50 Matches: 422
Percent Similarity: 81.40% Conservative: 33
Best Local Similarity: 75.49% Mismatches: 94
Query Match: 70.68% Indels: 11
DB: 3 Gaps: 7

US-09-787-016a-3 (1-562) x AAZ90579 (1-2867)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 232 ATGATGATATAAGGCGACCTGAGCATGAGGAGCACCACCAAGGCTATCAAAACCCACAGT 291

QY 21 LysGluPheArgLysThrTyrPheArgArgThrThrIleAlaLysArgGluGlyAla 40
Db 292 AAGGAGTTTCAGAAACCTGGGGTTTTCGAAAGAACCCAGATTCGCCAAACGTCGAGGGTGA 351

QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 352 GGAGACACGAGGCGGACCCAGTCAGTCAGCAACCA-----CAGCAGCATAACTCTCC 405

QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 406 CTGCGCCGACGTGACGCGCAACCAAAAGTACTGAGAGGGTAGAAGATTTCTTACACAG 465

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 466 GTTCGGCGCCGAGGAAAGAAATGTCCGGTGTCCCTGGAGATTCAGTCAGTCAGGCCACA 525

QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 526 TCTTCCACAGTCACCTGATGTGGAGACAGCTCCGAGGGGAGCGTTGAAAGCAGATTCAG 585

QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
Db 586 ATCAGAGTGGCCCTGTATCTGACTCTCTTAGGG-----AAAGAACATCTCCCTCTCTGA 642

QY 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 643 AAGCCAAAGAGAGGTGAAGAGGAGAGACACCTCTGACAGTCAGTCAGTCAGTCAGTCC 702

QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
Db 703 TTGAAGGAACCTCAGAACCGCTTCGGAGAAACGAGAGCAAGAACCTCTGTGAGAGGTCC 762

QY 181 LeuLysGlyLysSerArgLeuArgLysArgGluGluGluGluGluGluGluGluThr 200
Db 763 CTGAGGCGAGTCAGATTCCTGAGGAGAGAGCGCAGAGAGAGAGATTCGCGGAACT 822

QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
Db 823 ---GGAGTGTCCAAATAGGCTGCGAGCAGGACAGACCTCTCTGTAGCAGGAGCGCT 879

QY 220 GluAsnAspGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
Db 880 GAGGCTAGTCAGGACACAGTGTCCAGTCAGACAGACAGATGACATAGAAATCAGTTGGA 939

QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
Db 940 GGGAGGCGACTCAGGGAATACAGAGGAAACCCCGAGGAGCGGCGCAACCCAAAGCCT 999

QY 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
Db 1000 GAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCAATTCGCCGCCAGCCTCAACAAC 1059

QY 280 ArgPheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIle 299
Db 1060 AGTTTATGATTCGCTGTGATCGGTGTGAGAGTGTTCATCGTGTGCTGCTGTGTTATT 1119

QY 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
Db 1120 TCTGAGCGCGAGCGGCTCTCTGGAAGGAACCGGGAAGACTACATCTGCCCAATTCG 1179

QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
Db 1180 ACCATTTTCAGTGCAGGATGAGACAAACCGTAGCGCCACCAATGAGCAGGACTCTGGG 1239

QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
Db 1240 TGCAGATCTGTGGGTGCTGATGGCAAGCTGCAAGCATAGGAGCAGTAGGAGCAGAG 1299

QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379

DB 1300 TCCGAGAGACCGGATTAAGGCTAGGATTTAGAGAGCGGACCAACCCGAGCGCAAG 1359
QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
DB 1360 AAAAACTCAAGATTTCCAGGCTGTCTGAGAGGCTCTGTGGTCTCTTAATGCAATTCGC 1419
QY 400 ProGlyCysCysHisValAlaGlnProLysSerValTyCysSerAsnAspCysIleLeu 419
DB 1420 CTTGGGTGTTCAGGTGTAGCACAGCTGACTCTGTGTATTCAGTAATGACATTCCTC 1479
QY 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
DB 1480 AAACACGACGACGACCTACCATGATTTCTAAGTTCAAGTAAGACAAACAAACACCC 1539
QY 440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
DB 1540 AAGGAAAGAGCTCAACAGCAAGCCAGAAAGTTCAAGTCTTCCAAATGCAAGTTCAGGTG 1599
QY 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysIleGluThrVal 479
DB 1600 GGGATTAATCTTCTGTGCACAGACATAGCTGTGAGAAAGGAAACCAACCATG 1659
QY 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
DB 1660 AAGAAA---GTGATGCTGCTTCCAGAGTGTAGACTTCTGGAGGAGGAGCTGTGAG 1716
QY 500 SerSerThrProSerTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThr 519
DB 1717 AGCAGCACACCATCTCTGGGCAAGTGCACCACTCAATGTGTGAAGCCAGAGAGCCA 1776
QY 520 AlaAlaPro-----SerProSerLeuLeuTyLysCysMetTyHisLeuGlyVal 536
DB 1777 GAGAGCCCACTGCACTCTGCGCCACCTATTGAGTAAATGTACGTATCACCCAAAGGCT 1836
QY 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
DB 1837 GCCTTCCAGGCGCCCTCCCATC-ATCTGGGTGGCT---GCCTGGGGCTGTCTAGGA 1888
RESULT 11
AAC77933
ID AAC77933 standard; cDNA; 1764 BP.
XX AC AAC77933;
XX AC AAC77933;
XX DT 08-FEB-2001 (first entry)
XX DE Human cancer associated gene sequence SEQ ID NO:327.
XX KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropenic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX OS Homo sapiens.
XX PN W020005350-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005882.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.
XX P-PSDB; ABA43724.
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer.
XX Claim 1; Page 880-881; 2352pp; English.
XX AAC77607 to AAC78448 encode the human cancer associated proteins given in
XX ABA43998 to ABA44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnerable; immunomodulator;
XX antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
XX dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
XX neutropenic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions; graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and ABA44240 represent sequences used in the exemplification of
XX the present invention
XX SQ Sequence 1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;
Alignment Scores:
Pred. No.: 2,22e-97 Length: 1764
Score: 1706.00 Matches: 317
Percent Similarity: 99.37% Conservative: 0
Best Local Similarity: 99.37% Mismatches: 1
Query Match: 57.08% Indels: 2
DB: 3 Gaps: 0
US-09-787-016A-3 (1-562) x AAC77933 (1-1764)
QY 245 AspileLysAspGluGluProGlyAsp-LeuGlyArgProLysProGluCysGluGly 264
DB 2 GACATCAAGATGAGAGGCTGTGAGACTTTGGCCGACCGAGCTGATGATGAGGGTTA 61
QY 264 rAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArgPheMetIleCy 284
DB 62 CGACCCCAAGCGCCCTGTATTGCAATTTGGCCGAGCTCACACACACAGGTTTATGATTG 121
QY 284 sCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArg 304
DB 122 CTGTGACCGCTGTGAAGATGTTTCATGGCGATTGTGGGCATTTCTGAGCTCGAGG 181
QY 304 YArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThrIleLeuGlnVa 324
DB 182 GAGGCTTTTGAAGAAGATGGGAGAGACTATATCTGCCCAACTGCACCATTTCTCAAGT 241
QY 324 IGlNAspGluThrHisSerGluThrAlaAspGlnGluAlaLysTrpAcqProGlyAs 344
DB 242 GCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAATGAGACCTCGAGA 301
QY 344 pAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAsp 364
DB 302 TGTGATGGCACCATTGTACAGTATAGGACAAATAGAGCAGAGAGTCTAGCGAAGACCA 361
QY 364 nGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyIleLysLysLeuLys 384
DB 362 AGGATAAAGGCTAGAAATGAGAAAGCTGCAATCAATGAGTGGCAAGAGAACTCAAGAT 421
QY 384 ePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCys 404

Db 422 CTTCCAGCCTGTGATAGAGCGGCTGTGTGCTCAAAATGATTTATGGCCCCGGGTGTGTCA 481
 Qy 404 sValAlaGlnProAspSerValTyrcysSerAsnAspCysileuLysHisAlaAla 424
 Db 482 CGTGGGGCA-CCGAGCTCGTGTACTGCACTATGACTGTATCTCAACACGCGCAGC 540
 Qy 424 aThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetly 444
 Db 541 GACAATGAAGTTTCTAGCTCAGGTAAAGAACACAGAACCCAAAGCCTTAAAGAAAAGATGA 600
 Qy 444 sMetLysProGluLysProSerLeuProLysCysGlyValadlnAlaGlyileLysileSe 464
 Db 601 GATGAGCCAGAGAGCCAGCTTTCGAAATGCGGTGCTCAGCAGGTATTTAAATCTC 660
 Qy 464 rSerValHisLysArgProAlaProGluLysLysGluThrThrValLysLysAlaValVa 484
 Db 661 TTCTGTGACAGACAGACAGCTCCAGAAAAAAGAGACACACAGTCAAGAGCAGTGGT 720
 Qy 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSe 504
 Db 721 GGTCCCTGTGCGGAGTGAAGCCTCGGGAAGGAGCGAGCTTGTGAGAGCAGCAGCGCTC 780
 Qy 504 rTPAlaSerAspHisAsnTyrsnAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 781 GTGGGCGAGCGATCACAAATTACATGCAATGAGTAAAGCCAGAAAGACTGTCTCTCCCTCGCC 840
 Qy 524 oSerLeuLeuTyrcysMetTyrcysMetTyrcysMetTyrcysMetTyrcysMetTyrcys 544
 Db 841 GTCACTGTGTATAATGTATGATATCATCCAGGGGTGGCTCTGTGACCCCTCCCGGTTTC 900
 Qy 544 rPheTTPileAlaIleProTTPAlaCysProGlyLeuGlyValAlaAlaLeuLys 562
 Db 901 TTCTGATAGACCATCCCTGGGCTGTCCAGACTGGAGATTGCGACTTTGTGT 955
 RESULT 12
 AAS86575
 ID AAS86575 standard; cDNA; 2986 BP.
 AC AAS86575;
 XX 13-FEB-2002 (first entry)
 DT DNA encoding novel human diagnostic protein #22379.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US008631.
 PP 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG22388.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 22379; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS54197-AAS94584 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,11e-82 Length: 2986
 Score: 1467.00 Matches: 294
 Percent Similarity: 86.73% Conservative: 0
 Best Local Similarity: 86.73% Mismatches: 1
 Query Match: 49.08% Indels: 44
 DB: Gaps: 1

US-09-787-016A-3 (1-562) x AAS86575 (1-2986)

Qy 235 GluSerLysLeuGluGlyLysAlaAlaGlnAspLysLysAspGluProGlyAspLeu 254
 Db 3 GAGATGATTTGAGGGAAGGCGGCTCAGGATCATCAAGATGAGGAGGCTGAGACTTG 62
 Qy 255 GlyArgProLysProGluCysGluGlyTyrcysProAsnAlaLeuTyrcysileCysArg 274
 Db 63 GCGCGACCGAGCCTCAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCACTTTCGCG 122
 Qy 275 GlnProHisAsnAsnArgPheMetileCysCysAspArgCysGluGluThrPheHisGly 294
 Db 123 CAGCCTCACACACACAGGTTTATGATTTCTGTGACCGCTGTGAAGAAATGGTTTCATGGC 182
 Qy 295 AspCysValGlyileSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrc 314
 Db 183 GATTGTGTGGCATTCTCGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTAT 242
 Qy 315 IleCysProAsnCysThrileLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
 Db 243 ATCTGCCCAACTGCACCACTTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCAGAT 302
 Qy 335 GlnGlnGluAlaLysTTPArgProGlyAspAlaAspGlyThrAspCysThrSerileGly 354
 Db 303 CAGCAGGAGCTTAATGAGAGACTGGAGATGCTGTATGGCACCGATTGTACAGTATAGGA 362
 Qy 355 ThrileGluGlnLysSerSerGluAspGlnGlyileLysGlyArgGlnGluLysAlaAla 374
 Db 363 ACAATAGAGCAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTAGAATTGAGAAGCTGCA 422
 Qy 375 AsnProSerGlyLysLysLysLeuLysilePheGlnPro----- 387
 Db 423 AATCCAGTGGCGAGAGAGAACTCAAGATCTTCCAGCTTGGTCCCGGGCTGTGCCACC 482
 Qy 387 ----- 387
 Db 483 CAGCTGCTGTCTCTGCGCAGGATTTGGAAATGTCTGTCTAGAAAGCATCTCTGCTTTC 542
 Qy 388 -----ValileGluAlaProGlyAlaSerLysCysile 398
 Db 543 ACACCTTTCGACTGCATTAGCTGCAAGGATAGAGGCGCTGTGGCTCCAAATGTATT 602

PR	14-AUG-2000;	2000US-0225268P
PR	14-AUG-2000;	2000US-0225270P
PR	14-AUG-2000;	2000US-0225274P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225279P
PR	18-AUG-2000;	2000US-0226279P
PR	22-AUG-2000;	2000US-0226681P
PR	22-AUG-2000;	2000US-0226888P
PR	22-AUG-2000;	2000US-0227182P
PR	23-AUG-2000;	2000US-0227009P
PR	30-AUG-2000;	2000US-0228294P
PR	01-SEP-2000;	2000US-0229287P
PR	01-SEP-2000;	2000US-0229343P
PR	01-SEP-2000;	2000US-0229344P
PR	01-SEP-2000;	2000US-0229345P
PR	05-SEP-2000;	2000US-0229509P
PR	06-SEP-2000;	2000US-0229513P
PR	06-SEP-2000;	2000US-0230437P
PR	06-SEP-2000;	2000US-0230438P
PR	08-SEP-2000;	2000US-0231242P
PR	08-SEP-2000;	2000US-0231243P
PR	08-SEP-2000;	2000US-0231244P
PR	08-SEP-2000;	2000US-0231413P
PR	08-SEP-2000;	2000US-0231414P
PR	08-SEP-2000;	2000US-0232080P
PR	12-SEP-2000;	2000US-0232081P
PR	12-SEP-2000;	2000US-0232168P
PR	14-SEP-2000;	2000US-0232397P
PR	14-SEP-2000;	2000US-0232397P
PR	14-SEP-2000;	2000US-0232398P
PR	14-SEP-2000;	2000US-0232399P
PR	14-SEP-2000;	2000US-0232401P
PR	14-SEP-2000;	2000US-0233063P
PR	14-SEP-2000;	2000US-0233064P
PR	14-SEP-2000;	2000US-0233065P
PR	21-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-0234274P
PR	25-SEP-2000;	2000US-0234597P
PR	25-SEP-2000;	2000US-0234597P
PR	26-SEP-2000;	2000US-0234584P
PR	27-SEP-2000;	2000US-0235834P
PR	27-SEP-2000;	2000US-0235836P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236368P
PR	29-SEP-2000;	2000US-0236369P
PR	29-SEP-2000;	2000US-0236370P
PR	29-SEP-2000;	2000US-0236502P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	02-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239933P
PR	13-OCT-2000;	2000US-0239937P
PR	20-OCT-2000;	2000US-0240960P
PR	20-OCT-2000;	2000US-0241221P
PR	20-OCT-2000;	2000US-0241785P
PR	20-OCT-2000;	2000US-0241786P
PR	20-OCT-2000;	2000US-0241787P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241809P
PR	01-NOV-2000;	2000US-0244617P
PR	01-NOV-2000;	2000US-0244617P
PR	08-NOV-2000;	2000US-0246474P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0248523P
PR	08-NOV-2000;	2000US-0248524P
PR	08-NOV-2000;	2000US-0248525P
PR	08-NOV-2000;	2000US-0248526P

AAK67051

XX



Figure 1

КВ
1987

УС
АА
ОС

НИИ ВАРТАС.

XX
XX
ZNY

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PR 04-FEB-2000; 2000US-0180622

PR 02-MAR-2000; 2000US-018635

PR 17-MAR-2000; 2000US-019007

PR I9=MAI=2000; Z0000S=0Z0331
PR 07=JUN=2000; Z0000TS=0Z0946

29-JUN-2000; 2000US-021513
30-JUN-2000; 2000US-021513

PR 07-JUL-2000: 2000US-021688

PR 11-JUL-2000; 2000US-021749

PR 26-JUL-2000; 2000US-022098

PR 14-AUG-2000; 2000US-022451

PR 14-AUG-2000; 2000US-02252

PK I4-AUG-2000; 2000US-VZ2326
PP 14-AUG-2000: 2000US-Q22526

08-NOV-2000; 2000US-0246527P.
08-NOV-2000; 2000US-0246528P.
08-NOV-2000; 2000US-0246532P.
08-NOV-2000; 2000US-0246609P.
08-NOV-2000; 2000US-0246610P.
08-NOV-2000; 2000US-0246611P.
08-NOV-2000; 2000US-0246613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
17-NOV-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
Disclosure; SEQ ID NO 21863; 3071pp + Sequence Listing; English.
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting the
nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAK82169
represent sequences used in the exemplification of the present invention
SQ Sequence 30626 BP; 7102 A; 7610 C; 7917 G; 8097 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.44e-51 Length: 30626

Score: 989.00 Matches: 318
Percent Similarity: 18.71% Conservative: 0
Best Local Similarity: 18.71% Mismatches: 2
Query Match: 33.09% Indels: 1382
DB: 4 Gaps: 3
US-09-787-016A-3 (1-562) x AAK67051 (1-30626)
QY 243 AlaGlnAspIleValGlyAspGluProGlyAspLeuGlyArgProLysProGluCysGlu 262
Db 2 GCTCAGGACATCAAGATGAGAGGCTGGAGACTTGGGCGGACCGAGCGCTGAATGTGAG 61
QY 263 GlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn----- 279
Db 62 GGTTAGGACCCCAACGCCCTGTATGCAATTTGCCGCCAGCTCTCACACACACAGGTGGTC 121
QY 279 ----- 279
Db 122 ATGTGTGGGCACCCACAGAAATCCAGGGTTGCAAAACAGAAATTTCCATTTTTCATT 181
QY 279 ----- 279
Db 182 TTAAGGGTTACAAAGCTGGGCTGGATGAGCGATTGTTGGAAGATGAACAAGTTGTC 241
QY 279 ----- 279
Db 242 AAAAAAGAAACATAGAAGTGGCTAAGGACTTCTTAGAGCCACTCTCCACTCTGGTAG 301
QY 279 ----- 279
Db 302 CCACTGGCCACACATGGCTGTTCCAGTCAAGTAGGAGGTGAATAGCGCCCAACCCAG 361
QY 279 ----- 279
Db 362 TCAGGTAGCCACGTGCGCAGTCAGTCGCTCTGTGTAGACGGCATGGAGGAGATGTTCOG 421
QY 279 ----- 279
Db 422 TCATCTCCGGGCTGTGTCTGAGACTGCTCTTAGACTGCCCTTGGGAGACATAGCAGAG 481
QY 279 ----- 279
Db 482 GGGCATGTGACAGCCGCAAGTGTCTTAGTGCAACATTTGGGGCAGCTCAAAAGATCTCTG 541
QY 279 ----- 279
Db 542 CTAGGACTGGGTGTACAGACGGTGATAGGAGCGCTGTGGCTTGGCGCTGGGATAAGGCA 601
QY 279 ----- 279
Db 602 TTGTCCAGTTTTTGGGGGGAGACTTCTGTAGGATATCTGTGTGGATGCTTAGAGTTTG 661
QY 279 ----- 279
Db 662 CTCCTAAATGCGCTACTCTCTTAAAAAAGCGTGAGATAAATTTGGTTGAATGTTAAATCA 721
QY 279 ----- 279
Db 722 AAGGAAGGATAGGAGCGTTCTTAATCTCTACTCTTCTGTATATTTTGCATTTTTCATG 781
QY 279 ----- 279
Db 782 AAGTTAAACCAATTTGCTTATTATATGTTGTAAGAGTGTCTGTCACTCAGCTCCCATTTG 841
QY 280 -----ArgPheMetIleCysCysAspArgCysGluGluTIP 291
Db 842 CTAATACGGTTATTTTATAATTTTCAGGTTTATGTTTGTGTGACCGCTGTGAGGATGG 901
QY 292 PheHisGlyAspCysValGlyIleSerGluAlaArgGlyValGluLeuGluArgAsnGly 311
Db 902 TTTTCATGGCGATTGTGTGGCAATTTCTAGGCTCGAGGAGGCTTTTGGAAAGCAATGGG 961
QY 312 GluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGlu 331

Db 962 GAAGACTATATCTGCCAAACTGCACCAATTCGCAAGTCGAGGATGAGACTCATTCAGAA 1021
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 332 ThrAlaaspGlnGlnGlnAlaValTnApgProGlyVaspaAlaaspGlyThrAapCysThr 351
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1022 ACCGAGATCAGCAGGAGAGCTTAATGAGAGCTCGAGATGCTGATGAGCACCAGATTGACA 1081
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 352 SerIleGlyThrIleGluGlnLysSerSerGluaspGlnGlyIleValGlyArgIleGlu 371
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1082 AGTATAGGAACTATAGAGCAGAGCTCTAGCGAGAGCAAGGATTAAGGGGTAGAAATTGAG 1141
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 372 LysAlaAlaAanProSerGlyLysValLysValLysValPheGlnProVal 388
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1142 AAAGCTGCAATCCAGTGGCCAGAGAACTCAAGATCTCCAGGCTGT-AAAGTGTG 1200
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 388 ----- 388
Db 1201 FCCGCTGAGACGAGCAATTCAGAGTCGGCCTCACTGGTTCAGTCTCGCTCTAGATAGAA 1260
Qy ----- 388
Db 1261 ATTCTCCAGGACGGGTGCTGCACCCATGTGGGGGGTGCCTCTCCGGGTTCCTGGCT 1320
Qy ----- 388
Db 1321 CACTCCACCTGGGTGAGGCTTCTAATTTTATTCTTACCTGTCTACCTCTCTACTAT 1380
Qy ----- 388
Db 1381 TAATGACTTAATGTTGAATCAGCCCAATCTTTGAGGATTTGGGTGTGTTACATGCTT 1440
Qy ----- 388
Db 1441 TGTATGCAATTAGCAGAAAGAAATGTTTCTCAACGATGGGAGTGGGGTCTCTCTGTGTC 1500
Qy ----- 388
Db 1501 ACCTGTGTCTCCAGCTGCCTCGTGGTCTGGTCTGGCTGCCTCGGGGCTGTGTGTTGA 1560
Qy ----- 388
Db 1561 GAGTGGTTCCTCTCTTGACAGCTGTGTGACACTTAATCTCTCCCTCGATGTTTATG 1620
Qy ----- 388
Db 1621 ACATTTCCACATGCTGTCACTTCTGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
Qy ----- 388
Db 1681 CCCACAGGCGAGGTGTGACTTGGCGTGTGTTTCACTCTGTCTCCCTCCAGGGTCCCGGGCT 1740
Qy ----- 388
Db 1741 TGTCCCCACCCAGCTGCTCTCTCTGGCAGGATTTGGAAATTTGCTGTGTCTAGAGCAT 1800
Qy ----- 388
Db 1801 CTCTGCTTCACTCTTGCACTGTGCAATAGCTGCAAGGTGAGGATGTCCGAGGCTGGGC 1860
Qy ----- 388
Db 1861 CGTGCACCTGGAGCGTCCCTTGGAGGAGGAGGAGGCTGAAGCTCATGCTCGCTGCT 1920
Qy ----- 388
Db 1921 ATCAGAGCCACATCTGGGCAATTCAGAGATGAATGGAGATGGGAATTCATGAGAAAGTC 1980
Qy ----- 388
Db 1981 ATTGGCTTATGCTCTTATTTTAAATATATTGATCTGTTTTTTTGTGCACTTGAGGCT 2040
Qy ----- 388

Db 2041 TGGCTCTGTTGCCAGGCTGGAGTTTGGTGGCACAGTCCAGCTCACTGCAGCCTCGATC 2100
Qy ----- 388
Db 2101 TCCTGGGCTTGAGTGTCTCCACCTTAGCTCCCAAGTAGTGGGACTACAGGACAC 2160
Qy ----- 388
Db 2161 ACCACCAAGCCCAACATCTTTTTTAATTTGTTCATTTGTCTTGACAATACCTGTGCAAAAC 2220
Qy ----- 388
Db 2221 CTAAATTTGTATCTCAATTAATAACTTAGATTTCTGTGGAGGAAACTGATTAATTAATAT 2280
Qy ----- 388
Db 2281 TTAATTTGAGAAATAAGTTTGGCTTAAGTATGTGTGAGGACATGTTCTGTGTGTGTTGA 2340
Qy ----- 388
Db 2341 GGCCTGATCATCGAGACAGACTTTGTGTAACAGCAGTGGTCAAGCTGTTTGTGGTGGTGA 2400
Qy ----- 388
Db 2401 GTTTCTCTCTGCTGACGATGAGTTACTGCAGTCTCTGGCTTGAAGGAGCTCACATTTATG 2460
Qy ----- 388
Db 2461 CCTCCATTTCCATGGGAAGCCTGGACCTGCTGCTCAGGTCTCACAGGACTGCAGTCAOG 2520
Qy ----- 388
Db 2521 TGCAGCGGGGCTGTGTCTCCAGAGCTTGGCAGGGGATTCCTTCCAGAGCCCTCAC 2580
Qy ----- 388
Db 2581 GCTGTTTTTACTATATAAGTGTGTCTGGCGGGTGTGTGGCTCACACCTGTCAATCCCA 2640
Qy ----- 388
Db 2641 GCACCTTGAGAAAGTCAAGATGGGAGGATCACTTGAGGCCAAGAGTTGAGACAGCCTGG 2700
Qy ----- 388
Db 2701 GCAAGCTAGTAGACCCCATCTCTACAAAAAATAATAATAGCTGGGCGTGGTGA 2760
Qy ----- 388
Db 2761 TGTGCTGTAGTCCAGCTACTGGGAGGCTGGGCTGGGAGATCACTTGAGCCAGGAC 2820
Qy ----- 388
Db 2821 ATCAAGGCTGCAGTGAAGTATGTTGCTCACTGCTCTCCGGCTGGGCAACAGAGGAG 2880
Qy ----- 388
Db 2881 ACCCTGTCTCAAAACAAAAAAGTATGTTTCATGCAATTTGGGGGGCATACTTACACT 2940
Qy ----- 388
Db 2941 GAABAAATTAATCTGTTTATCTGACATTCGATTTCAATTGGCATACTGTATTTTACTTG 3000
Qy ----- 388
Db 3001 GCAGCTTGGCCTCAGATCTTAGCCACAGCTAATCTCTCAAGCAGGAGGATTTACTTC 3060
Qy ----- 388
Db 3061 AAACCCAGCTATGAAGTCTTCTCTCTTCATGAAGAGCCAGTTCCTTTTAAAGGGTTTT 3120
Qy ----- 388
Db 3121 TACCTGATTTAGCGGGAGCCCAATTTGACTTGCAGTCAATCCAGTTAGATCCTTAATTACAT 3180

Search completed: May 1, 2004, 17:45:21
Job time : 611 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: April 30, 2004, 08:54:02 ; Search time 50 Seconds
(without alignments)
3115.611 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGDPSNEEAPKAIKPTS.....RSFWIAIPWACGLGVAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1138120 segs, 277189581 residues
Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	869	29.1	181	9	US-09-925-301-1169
2	181.5	6.1	288	9	US-09-764-864-1258
3	174.5	5.8	669	10	US-09-823-187-88
4	174.5	5.8	669	10	US-09-863-776-53
5	169	5.7	676	10	US-09-823-187-24
6	169	5.7	676	10	US-09-863-776-14
7	169	5.7	676	14	US-10-203-708-28
8	169	5.7	678	10	US-09-823-187-87
9	169	5.7	678	10	US-09-863-776-52
10	169	5.7	717	14	US-10-203-708-29
11	167	5.6	667	10	US-09-863-776-16
12	167	5.6	667	10	US-09-863-776-18
13	166	5.6	555	14	US-10-106-698-5847
14	166	5.6	670	10	US-09-823-187-86
15	166	5.6	670	10	US-09-863-776-51

Sequence 308, App	16	166	5.6	671	10	US-09-946-374-308	Sequence 85, Appl
Sequence 20, Appl	17	166	5.6	671	10	US-09-823-187-85	Sequence 20, Appl
Sequence 346, App	18	166	5.6	671	10	US-09-863-776-20	Sequence 346, App
Sequence 346, App	19	166	5.6	671	12	US-10-206-915-346	Sequence 346, App
Sequence 346, App	20	166	5.6	671	12	US-10-199-670-346	Sequence 346, App
Sequence 346, App	21	166	5.6	671	12	US-10-201-858-346	Sequence 346, App
Sequence 346, App	22	166	5.6	671	12	US-10-205-890-346	Sequence 346, App
Sequence 346, App	23	166	5.6	671	12	US-10-208-024-346	Sequence 346, App
Sequence 346, App	24	166	5.6	671	12	US-10-201-853-346	Sequence 346, App
Sequence 346, App	25	166	5.6	671	12	US-10-174-581-346	Sequence 346, App
Sequence 346, App	26	166	5.6	671	12	US-10-176-483-346	Sequence 346, App
Sequence 346, App	27	166	5.6	671	12	US-10-176-749-346	Sequence 346, App
Sequence 346, App	28	166	5.6	671	12	US-10-176-914-346	Sequence 346, App
Sequence 346, App	29	166	5.6	671	12	US-10-176-915-346	Sequence 346, App
Sequence 308, App	30	166	5.6	671	12	US-10-006-485A-308	Sequence 308, App
Sequence 308, App	31	166	5.6	671	12	US-10-013-907A-308	Sequence 308, App
Sequence 346, App	32	166	5.6	671	12	US-10-015-499A-308	Sequence 346, App
Sequence 346, App	33	166	5.6	671	12	US-10-176-484-346	Sequence 346, App
Sequence 346, App	34	166	5.6	671	12	US-10-180-550-346	Sequence 346, App
Sequence 346, App	35	166	5.6	671	12	US-10-183-014-346	Sequence 346, App
Sequence 346, App	36	166	5.6	671	12	US-10-187-738-346	Sequence 346, App
Sequence 346, App	37	166	5.6	671	12	US-10-187-740-346	Sequence 346, App
Sequence 346, App	38	166	5.6	671	12	US-10-187-883-346	Sequence 346, App
Sequence 346, App	39	166	5.6	671	12	US-10-194-363-346	Sequence 346, App
Sequence 346, App	40	166	5.6	671	12	US-10-194-460-346	Sequence 346, App
Sequence 346, App	41	166	5.6	671	12	US-10-194-463-346	Sequence 346, App
Sequence 346, App	42	166	5.6	671	12	US-10-194-484-346	Sequence 346, App
Sequence 346, App	43	166	5.6	671	12	US-10-195-884-346	Sequence 346, App
Sequence 346, App	44	166	5.6	671	12	US-10-195-896-346	Sequence 346, App
Sequence 346, App	45	166	5.6	671	12	US-10-196-744-346	Sequence 346, App

ALIGNMENTS

RESULT 1
US-09-925-301-1169
; Sequence 1169, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1169
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1169

Query Match	29.1%;	Score	869;	DB	9;	Length	181;
Best Local Similarity	99.4%;	Pred. No.	6.8e-54;				
Matches	154;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						
Qy	254	LCRPKPECEGYPNALYCLICRQPHNNRPMICCDRCBFWFHGDCVGISEARGRLLENGED	313				
Db	10	LCRPKPECEGYPNALYCLICRQPHNNRPMICCDRCBFWFHGDCVGISEARGRLLENGED	69				
Qy	314	YICPNCITLQVDETHSETADQOEAKWRPGDADGTCTSIGTIEQKSSDQGIKGRIEKA	373				
Db	70	YICPNCITLQVDETHSETADQOEAKWRPGDADGTCTSIGTIEQKSSDQGIKGRIEKA	129				
Qy	374	ANPSGKKLKIPOPVLEAFGASKICPGCHVAQP	408				
Db	130	ANPSGKKLKIPOPVLEAFGASKICPGCHVAQP	164				

; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-24

Query Match
5.7%; Score 169; DB 10; Length 676;
Best Local Similarity 20.4%; Pred. No. 0.0018;
Matches 115; Conservative 78; Mismatches 209; Indels 162; Gaps 23;

QY 1 MDDKGDPSNEBAPKAIPKTSKEFRKTWGFRTTIAKREGADAEADPLEPPPPQOQLGLS 60
DB 140 MESDSDSKSDNSGLK-----RKTAL-KVSVSKARKASSDLDOASVSPSEENSES 192
QY 61 LRRSGRQPKRTVEQFLTI-ARRG-----RRSMPVSLSDSGEPTSC-----PATDAE 108
DB 193 SSESEKTSQDQFTPEKKAAPRRGPKGKPKPAEKPLPKPRGKPKPPEPPSS-----SS 302
QY 109 TASGSESVESAS-----ETRSQFQSASTAVK-----ERPASSEKVKGGDDHDDT 151
DB 253 SASSSSSSSSSSDSDSVSVKPPGKPKPAEKPLPKPRGKPKPPEPPSS-----SS 302
QY 152 SDSDSGL-TLKEIQNR-----LRRKREOFTPEPLKIQSLAKKRE---EGPA 198
DB 303 SDSDSDEVDRISEWKRRDEARRRLEARRRREBEELRRLEQKEEKKERRERADRGEA 362
QY 199 E--TVGSEASDTVEGVLPKQEPENDQGVWSQAGKD-DRESKLEKGAQDINDPEPGDLG 255
DB 363 ERGSGSSGDELREDDPEVKKEGRGKRGPPSSSDSPEAELEAKKSAK-----414
QY 256 RPKPCBGYDNPALYICIQPHNNRFTMCCDCEHFWGDCVGISEARGRLLENGEYI 315
DB 415 --KQSSSTEP-----ARKPG-----428
QY 316 CPNCTILQVDETHSETAQQAQKWRPGDADQDCTSIGTIBQKSSDQGIKRIEKAAN 375
DB 429 -----QKERVPRPEKQA--RPVKVET-----RKSEGFSDNRKVEKKKE 468
QY 376 PSKKKL-----IPQVIEAPGASKIGPCCHVAQPSVCSNDCLIKH--AAATWK 427
DB 469 PSVBEKQLKLSHSEIKFALKVDSFDVKRCILNA-----LEELGLTQVTSQILQKNTDVVATLK 524
QY 428 FL---SSGKEQKPKPKKMKPKPEPSLPKCGAQAQGIKISSVHKRPAPKPKETTAKAV 484
DB 525 KIRRYKANKDVMEKAAEVTVLRSVLPKIEAVKVKVAGKEKKEKAKEL-----575
QY 485 VPARGSEALGKEAACSSSTPSWASD 508
DB 576 ---AGEELAGEELAGEAPQEAED 597

RESULT 6

US-09-863-776-14
; Sequence 14, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytex, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K

; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Esha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863,776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206,679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207,798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208,263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208,831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209,451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210,060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219,507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221,337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221,927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263,135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263,694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-863-776-14

Query Match

5.7%; Score 169; DB 10; Length 676;
Best Local Similarity 20.4%; Pred. No. 0.0018;
Matches 115; Conservative 78; Mismatches 209; Indels 162; Gaps 23;

QY 1 MDDKGDPSNEBAPKAIPKTSKEFRKTWGFRTTIAKREGADAEADPLEPPPPQOQLGLS 60
DB 140 MESDSDSKSDNSGLK-----RKTAL-KVSVSKARKASSDLDOASVSPSEENSES 192
QY 61 LRRSGRQPKRTVEQFLTI-ARRG-----RRSMPVSLSDSGEPTSC-----PATDAE 108
DB 193 SSESEKTSQDQFTPEKKAAPRRGPKGKPKPAEKPLPKPRGKPKPPEPPSS-----SS 302
QY 109 TASGSESVESAS-----ETRSQFQSASTAVK-----ERPASSEKVKGGDDHDDT 151
DB 253 SASSSSSSSSSSDSDSVSVKPPGKPKPAEKPLPKPRGKPKPPEPPSS-----SS 302
QY 152 SDSDSGL-TLKEIQNR-----LRRKREOFTPEPLKIQSLAKKRE---EGPA 198
DB 303 SDSDSDEVDRISEWKRRDEARRRLEARRRREBEELRRLEQKEEKKERRERADRGEA 362
QY 199 E--TVGSEASDTVEGVLPKQEPENDQGVWSQAGKD-DRESKLEKGAQDINDPEPGDLG 255
DB 363 ERGSGSSGDELREDDPEVKKEGRGKRGPPSSSDSPEAELEAKKSAK-----414

QY 256 RPKPCEGVDNALYICICQPHNNRPMICCDRCBFWHGDGCVGISBARGRLRNGDYI 315
DB 415 --KPSSTP-----AKPG-----428
QY 316 CPNCTILQVDETHSETAQOAKWRPGDADGTDCTISGTHOKSSEDOGINKRLEKKAAN 375
DB 429 -----QKRVPRPEKQAK--PVKVERT-----RKSEGFSDMRKVKKEKE 468
QY 376 PSKKKLLK-----IFQPVIEAPGASKICIGPGCHVAQPDSDVYCSNDCILKH--AAATMK 427
DB 469 PSVEKLLQKLHSEIKFALXVSDPVKRCNA-----LEBLGTQVTSQILQKNTDVVATLK 524
QY 428 FL-----SSGKEQKPKPKKMKPKPSLPCGAQAQIKLSSVHKRPAPKPKETTVAUV 484
DB 525 KIRRYKANKDVMEKAAEVYTRLSRVLPKIEAVQKVNKAGMEKEKAEKL-----575
QY 485 VPASSEALGKEACSSSTPSWASD 508
DB 576 ---AGBELAGEELAGEAPQEKAD 597

RESULT 7
US-10-203-708-28
Sequence 28, Application US/10203708
Publication No. US20030149238A1
GENERAL INFORMATION:
APPLICANT: SMITHKLINE BEECHAM CORPORATION
APPLICANT: SMITHKLINE BEECHAM P.L.C.
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP50013
CURRENT APPLICATION NUMBER: US/10/203,708
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: PCT/US01/04703
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,172
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 60/186,084
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 676
TYPE: PRT
ORGANISM: Homo sapiens
US-10-203-708-28

Query Match 5.7%; Score 169; DB 14; Length 676;
Best Local Similarity 20.2%; Pred. No. 0.0016;
Matches 115; Conservative 78; Mismatches 200; Indels 176; Gaps 23;
QY 2 DDGKDPSEBAPKAIPKTSKFRKTWGFRTT-----IAKREGAGDAEADPLEPPPPQ 55
DB 143 DSDSDKSSDMS-----GLKRTPALXVSVSKRARKASSDLDQASVSPSE 187
QY 56 QLGLSLRRSGQPKRTERVQFUTL-ARRG-----RRMPVLEDSGFTSC-----P 103
DB 188 ENESSSESEKTSQDPTPEKAAVAPRPGPLGGRKKKXKAPSASDSDSKADSDGAKPEP 247
QY 104 ATDAETASEGVSAS-----ETESGQSASTAVK-----ERPASKEVKGGD 146
DB 248 VAMARSASSSSSSSDSDSVSVKPKGRKPAKPLPKPRGKPKPRPSS-----300
QY 147 DHDITSDSDGL--TLKELQNR-----LRKREQRTPEPLKIGIQLRKKRRE-- 194
DB 301 ---SSSDSDSDVDRISEWKRDEARRELEARRRREQBELRLRQEKERERRERA 357
QY 195 -EGPAP--TVGSEASDTEGVLPSQBPENDGVVSOAGKD--DEESKLEKGAQDIDEE 250
DB 358 DRGEAERGSGSGSDELREDDPEVKKGRGKRGPPSSSDSEPEASLEERAKKSAK---414
QY 251 PGDLGRPKPECEGYDYNALYICICQPHNNRPMICCDRCBFWHGDGCVGISBARGRLERN 310
DB 415 -----KPSSTP-----AKPG-----428

QY 311 GEDVICPNCTILQVDETHSETAQOAKWRPGDADGTDCTISGTHOKSSEDOGINKR 370
DB 429 -----QKRVPRPEKQAK--PVKVERT-----RKSEGFSDMRKVK 463
QY 371 EKAANPSKKKLLK-----IFQPVIEAPGASKICIGPGCHVAQPDSDVYCSNDCILKH--A 422
DB 464 EKKSEPSVEKLLQKLHSEIKFALXVSDPVKRCNA-----LEBLGTQVTSQILQKNTDV 519
QY 423 AATMKFL-----SSGKEQKPKPKKMKPKPSLPCGAQAQIKLSSVHKRPAPKPKETT 479
DB 520 VATLKIRRYKANKDVMEKAAEVYTRLSRVLPKIEAVQKVNKAGMEKEKAEKL-----575
QY 480 KKAUVVPASSEALGKEACSSSTPSWASD 508
DB 576 -----AGBELAGEELAGEAPQEKAD 597

RESULT 8
US-09-823-187-87
Sequence 87, Application US/09823187
Publication No. US20030096952A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumdar, Kumud
APPLICANT: Padigaru, Muralidhar
APPLICANT: Patturajan, Meera
APPLICANT: Shimmets, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 87
LENGTH: 678
TYPE: PRT
ORGANISM: Mus musculus
US-09-823-187-87

Query Match 5.7%; Score 169; DB 10; Length 678;
Best Local Similarity 20.1%; Pred. No. 0.0016;
Matches 132; Conservative 80; Mismatches 182; Indels 264; Gaps 32;
QY 2 DDGKDPSEBAPKAIPKTSKFRKTWGFRTT-----IAKREGAGDAEADPLEPPPPQ 55
DB 143 DSDSDKSSDMS-----GLKRTPALXVSVSKRARKASSDLDQASVSPSE 187

```
QY 56 QLGSLRSRGRQKRTERVE-----QFLTIAR--RGRSRMPVSLDSGPT--SCPA- 104
D 188 -----DSEPSSEKTDQFTPEKTAAPRRG-----PLGGRKKKHTGVACPK 235
QY 105 -----TDAETASGVSVE-----SASETSQPSASTAVK----- 133
D 236 VPSASDSDSKADSGAKKEPVVTAQPSPPSSSSSSSSSDSDSVVKKPPGRKPAKPP 295
QY 134 -----ERPASSEKVGKGDHDDTSDSDGLTKELQ-----RLRKK 171
D 296 PKPRGRPKPPEPST-----SSSDSDSGEVDRISWKRRDERRERLEARR 345
QY 172 REQEPTEPLKGIQSLRKKRREBEPVTAQPSPPSSSSSSSDSDSVVKKPPGRKPAKPP 231
D 346 REQEEELRLRE-QEREKERRKE-RAERCGS-----SGELEDEBPV----- 386
QY 232 DDRESKLGKAAQDIKDEEP-GDLGR-----PKPCEGYDPNALYICROPHNNRPMIC 284
D 387 KKSRRKARGTPTSSSDSEPEGLKGGKLAQKSLPGSES-----ARKP----- 432
QY 285 CDRCEWFHGDVGISEARGLLERNGEDYICPNCITLOVDTHSETADQQAQKVRPGD 344
D 433 -----GQEKRR-----YCSNDCILKHA-----A 424
QY 345 ADGTDCTSIGTIRKSSSDGQIKRIEKAANPSGKKKX-----IFQVIEAPGASKCI 398
D 452 VERT-----RKSEGLSLERKKEKKEPSVEERLQKLSHSEIKFALKVNDPVRKCL 502
QY 399 GP-----GCCHVA-----QDSV-----YCSNDCILKHA-----A 424
D 503 SALEBELGTQVTSQILQKNTDVVATLKKIRRYKANDVMKAAAEVYTRLSRVLGPKVEA 562
QY 425 TMKFLSSGKEQKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKK 481
D 563 LQKVNKAGAKERADNEKLEEQ-----GEOAPRELAEDPSTRAPVNGEATSQK 614
QY 482 AVVVPARSEALGKEACESTPWSADH-----NNAVKP-----EKTAAPS 525
D 615 GENMEDAQDGDGGS--EDPGSGSEELHSDPRDSDPAKPNRQHERTRLASES 670

RESULT 9
US-09-863-776-52
; Sequence 52, Application US/09863776
; Publication NO. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderne, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Edna
; TITLE OF INVENTION: No. US20030198953A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863, 776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540, 763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206, 679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207, 748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207, 798
; PRIOR FILING DATE: 2000-05-30
```

```
; PRIOR APPLICATION NUMBER: 60/208, 263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208, 831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209, 451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210, 060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219, 507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221, 337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221, 927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263, 135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263, 688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263, 694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 52
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-863-776-52

Query Match 5.78; Score 169; DB 10; Length 678;
Best Local Similarity 20.1%; Pred. No. 0.00187;
Matches 132; Conservative 80; Mismatches 182; Indels 264; Gaps 32;

QY 2 DDGDDPNEAPKAIPKTSKFRKTWGRFTT-----TAKREGAGDAEADPLEPPPOQ 55
D 143 DSDSDKSSDHS-----GLKKTPLVKVSVSKRARRASSDLDQASVSPSEE 187
QY 56 QLGSLRSRGRQKRTERVE-----QFLTIAR--RGRSRMPVSLDSGPT--SCPA- 104
D 188 -----DSEPSSEKTDQFTPEKTAAPRRG-----PLGGRKKKHTGVACPK 235
QY 105 -----TDAETASGVSVE-----SASETSQPSASTAVK----- 133
D 236 VPSASDSDSKADSGAKKEPVVTAQPSPPSSSSSSSSSDSDSVVKKPPGRKPAKPP 295
QY 134 -----ERPASSEKVGKGDHDDTSDSDGLTKELQ-----RLRKK 171
D 296 PKPRGRPKPPEPST-----SSSDSDSGEVDRISWKRRDERRERLEARR 345
QY 172 REQEPTEPLKGIQSLRKKRREBEPVTAQPSPPSSSSSSSDSDSVVKKPPGRKPAKPP 231
D 346 REQEEELRLRE-QEREKERRKE-RAERCGS-----SGELEDEBPV----- 386
QY 232 DDRESKLGKAAQDIKDEEP-GDLGR-----PKPCEGYDPNALYICROPHNNRPMIC 284
D 387 KKSRRKARGTPTSSSDSEPEGLKGGKLAQKSLPGSES-----ARKP----- 432
QY 285 CDRCEWFHGDVGISEARGLLERNGEDYICPNCITLOVDTHSETADQQAQKVRPGD 344
D 433 -----GQEKRR-----YCSNDCILKHA-----A 424
QY 345 ADGTDCTSIGTIRKSSSDGQIKRIEKAANPSGKKKX-----IFQVIEAPGASKCI 398
D 452 VERT-----RKSEGLSLERKKEKKEPSVEERLQKLSHSEIKFALKVNDPVRKCL 502
QY 399 GP-----GCCHVA-----QDSV-----YCSNDCILKHA-----A 424
D 503 SALEBELGTQVTSQILQKNTDVVATLKKIRRYKANDVMKAAAEVYTRLSRVLGPKVEA 562
QY 425 TMKFLSSGKEQKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKK 481
D 563 LQKVNKAGAKERADNEKLEEQ-----GEOAPRELAEDPSTRAPVNGEATSQK 614
QY 482 AVVVPARSEALGKEACESTPWSADH-----NNAVKP-----EKTAAPS 525
```

Db 615 GENMEDAQRDQDS--EDGPRGSSSELHSDSPRDNSDPAKPGNVERQDHERTRLASES 670

RESULT 10

US-10-203-708-29

Sequence 29, Application US/10203708

Publication No. US20030149238A1

GENERAL INFORMATION:

APPLICANT: SMITHKLINE BEECHAM CORPORATION

APPLICANT: SMITHKLINE BEECHAM P.L.C.

FILE OF INVENTION: NOVEL COMPOUNDS

FILE REFERENCE: GP50013

CURRENT APPLICATION NUMBER: US/10/203.708

CURRENT FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: PCT/US01/04703

PRIOR FILING DATE: 2001-02-14

PRIOR APPLICATION NUMBER: 60/182,172

PRIOR FILING DATE: 2000-02-14

PRIOR APPLICATION NUMBER: 60/186,084

PRIOR FILING DATE: 2000-02-29

NUMBER OF SEQ ID NOS: 46

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 29

LENGTH: 717

TYPE: PRT

ORGANISM: Homo sapiens

US-10-203-708-29

Query Match 5.7%; Score 169; DB 14; Length 717;

Best Local Similarity 20.2%; Pred. No. 0.0019;

Matches 115, Conservative 78; Mismatches 200; Indels 176; Gaps 23;

Qy 2 DDKGDPNBEAPKAIKTSIKFTWGRFTT-----IAKREGAGDAEADPLEPPPPQ 55

Db 184 DSDSDKSSDMS-----GLKRTPALKMSVSKRAKASDLDQASVSPSE 228

Qy 56 QLGLSLRSGRQPKTERVEQFLTI-ARRRG-----RRSMFVSLSDSGEPTSC--P 103

Db 229 ENSESSSEKTSQDFTPEKAAVAPRPRGPGICGRKKKAPASDSKADSDGAKPEP 288

Qy 104 ATDAETASBGSVESAS-----ETSGQSASTAVK-----ERPASSEKVGGD 146

Db 289 VAMARSASSSSSSSSSDSVSVKPRGRKPAEKPLPKPRGRKPKPRPPSS----- 341

Qy 147 DHDTSDSDGL-TLXLQNR-----LREKREQEPTEPLKGIQSLRKKRE-- 194

Db 342 ---SSSDSDSDVDRISEWKRDEARRELEARRRQSEELRLRQEKKEKRRERA 398

Qy 195 -EGPAB--FVSGEASDTVEGLPSQBPENDQGVVVSQAGKD-DRESKLEGKAAQDIKDE 250

Db 399 DRGEARGSGSGDELREDDPEPVKGRKGRGRGPPSSSDSEPEALEAKKSASAK--- 455

Qy 251 PGDLGRPKPECEGYDPNALYICIQPHNNRPMICCDRCSEWFHGDGCVISEARGLLERN 310

Db 456 -----KQSSSTEP-----ARKPG----- 469

Qy 311 GEDYICPNTLLQVDETHSETAQQAQKWPAGDADGDTCTSIGTIBQKSSDQIGIRI 370

Db 470 -----QKEKRVPEEKQQAQ--PVKVERT-----RKSESGFSDMRKV 504

Qy 371 EKANPSSGKKLK-----IPQVIEPAGSKCTIGPCCHVAQSDSVYCSNDILKH--A 422

Db 505 EKKKEPSVEEKQLKHSIEIPALKVDSPDKRCINA-----LEEIGTLQVTSQILQKNTDV 560

Qy 423 AATMKFL---SSGKEQKPKPKKMKPKPKPKPKKPKKPKKPKKPKKPKKPKKPKKPK 479

Db 561 VATLUKKIRYKANDVMEKAEVYTRLSRVLPKIEAVQKNCAGMEKEKAEKL----- 616

Qy 480 KCAVVVPARSALGKBAACESSTSNASD 508

Db 617 -----AGEELAGEEAPQEKAD 638

RESULT 11

US-09-863-776-16

Sequence 16, Application US/09863776

Publication No. US20030198953A1

GENERAL INFORMATION:

APPLICANT: Splytek, Kimberly A

APPLICANT: Majumder, Rumud

APPLICANT: Tchurnev, Velizar T

APPLICANT: Mishra, Vishnu

APPLICANT: Padigaru, Muralidhara

APPLICANT: Spaderna, Steven K

APPLICANT: Shenoy, Suresh K

APPLICANT: Rastelli, Luca

APPLICANT: Li, Li

APPLICANT: Taupier, Raymond J

APPLICANT: Gangolli, Esha

FILE OF INVENTION: NO. US20030198953A1e1 Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-020

CURRENT APPLICATION NUMBER: US/09/863,776

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 09/540,763

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/206,679

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,688

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,829

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/207,748

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/207,798

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/208,263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 60/208,831

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/209,451

PRIOR FILING DATE: 2000-06-05

PRIOR APPLICATION NUMBER: 60/210,060

PRIOR FILING DATE: 2000-06-07

PRIOR APPLICATION NUMBER: 60/219,507

PRIOR FILING DATE: 2000-07-20

PRIOR APPLICATION NUMBER: 60/221,337

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/221,927

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: 60/263,135

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: 60/263,688

PRIOR FILING DATE: 2001-01-24

PRIOR APPLICATION NUMBER: 60/263,694

PRIOR FILING DATE: 2001-01-24

NUMBER OF SEQ ID NOS: 155

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 16

LENGTH: 667

TYPE: PRT

ORGANISM: Homo sapiens

US-09-863-776-16

Query Match 5.6%; Score 167; DB 10; Length 667;

Best Local Similarity 20.5%; Pred. No. 0.0024;

Matches 115, Conservative 78; Mismatches 208; Indels 160; Gaps 23;

Qy 1 MDKGGPDSNEAPKAIKTSIKFTWGRFTTIAKREGAGDAEADPLEPPPPQQLGLS 60

Db 140 MESDSDSDKSSDMSGLK-----RKTPLAL-KVSVSKRAKASDLDQASVSPSEENSES 192

Qy 61 LRPSGRQPKTERVEQFLTI-ARRRG-----RRSMFVSLSDSGEPTSC-----PATDAE 108

Db 193 SRSRKTSDQDFTPEKAAVAPRPRGPGICGRKKKAPASDSKADSDGAKPEPVAMAR 252


```

Db      520  VATLKKIRRYKANKDMVEKAAEVYTRLKSRVLGPKIEAVQKVKACGMEKAKBEKL----- 575

Qy      480  KXAVVPARFSEALGKEAACESSTPSWASD 508

Db      576  -----AGEELAGEEAPQEKADKPSTD 597

```

RESULT 15

```

US-09-863-776-51
; Sequence 51, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Ganggalli, Zsha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863,776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206,679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207,798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208,263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208,831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209,451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210,060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219,507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221,337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221,927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263,135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263,694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-863-776-51

```

Query Match 5.6%; Score 166; DB 10; Length 670;
Best Local Similarity 20.0%; Pred. No. 0.0029;
Matches 114; Conservative 80; Mismatches 199; Indels 176; Gaps 23;

```

143 DSDSDKSDNS -----GLKRTPTALXSVSKRARKASDLDLOASVSPSEE 187
Qy 56 QLGSLRSRGRQPKRTERVQFLTI-ARRRG-----RRSMFVSLSDSGEPTSC-----P 103
Db 188 ENSESSSESEKTSQDQDTPFEKKAIVRAPRGRGLGRKKKAPASDSDSKADSDGAKPEP 247
Qy 104 ATTAETASGSESVAS -----ETFSGQSASTAK-----EPASSEKVKGD 146
Db 248 VAMARSASSSSSSSSSDSDSVSVKPRGRKPAKPLPKPRGRKPKPRFPSS ----- 300
Qy 147 DHDTSDDSDGL-TLXLQNR -----LRKRQEPTERPLKGTQSLRKKRRE----- 194
Db 301 ---SSSDSDSDVDRISEMKRRDEARRELEARRRQEELRLRQEKEKEKERRA 357
Qy 195 -EGPAE---TWGSEASDFVEGVLPSKQBPENDGVVSQAGD-DRESKLEGKAAQIDKBE 250
Db 358 DRGEAREGGSSGDELRDEDDPEVKRGRKGRGPPSSSSSEPEALEREAKKSAAK--- 414
Qy 251 PGDLGRPKCEBGVDPNALXYCICRQPHNNRFMICCDRCBWFHGDVCVGISBARGLLERN 310
Db 415 -----KQSSSTEP-----ARKPG----- 428
Qy 311 GEDYICNCTILQVODETHSETADQAKNRPGDADGTDCTSIGTIEOKSSDQIKGRI 370
Db 429 -----QKGRVREKQAK--PVKVERT-----RKSEGFSDMRKV 463
Qy 371 EKAANPSGKKLX-----IFQVIEAPGASKICIGPCCHVAQPDVSVVSCNDILKH--A 422
Db 464 EKKKEPSEBKLQKLHSEIKFALKVDPDVKRCINA---LEELGTLQVTSQILQKNTDV 519
Qy 423 AATMKEL---SSGKBQKPKKPKKMKPEKPSLPGCGAQAQIKLSSVHKRPAPKKETT 479
Db 520 VAILKKIRRYKANDVMEKAAEVVTRUKSRVLGPKIEBAVKVKNKAGMEKKBKBEKL----- 575
Qy 480 KGAVVVPARSEALGKEACESSSTPSWASD 508
Db 576 -----AGEELAGERAPOKAEDKPDST 597

```

Search completed: April 30, 2004, 08:57:36
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:52:47 ; Search time 22 Seconds
(without alignments)
1318.809 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKDPSENEAPRAIKPTS.....RSFWIAIPWACPGIGVAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCITUS-COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174.5	5.8	2289	3	US-09-051-019-2 Sequence 2, Appli
2	164.5	5.5	1075	4	US-09-252-991A-18387 Sequence 18387, A
3	161.5	5.4	1780	1	US-08-769-309A-5 Sequence 5, Appli
4	161.5	5.4	1780	3	US-08-994-570-5 Sequence 5, Appli
5	161.5	5.4	238	4	US-09-257-179-80 Sequence 80, Appl
6	158.5	5.3	779	4	US-10-164-595-56 Sequence 56, Appl
7	157	5.3	1404	4	US-08-801-308-1 Sequence 1, Appli
8	152.5	5.1	735	4	US-10-164-595-80 Sequence 80, Appl
9	152.5	5.1	784	4	US-10-164-595-79 Sequence 79, Appl
10	152.5	5.1	843	4	US-10-164-595-54 Sequence 54, Appl
11	152.5	5.1	1596	4	US-08-978-277A-4 Sequence 4, Appli
12	149.5	5.0	2468	4	US-09-976-594-726 Sequence 726, App
13	149.5	5.0	8991	1	US-08-714-741-32 Sequence 32, Appl
14	148.5	5.0	1805	1	US-07-853-913-2 Sequence 2, Appli
15	148.5	4.9	1618	1	US-07-853-913-4 Sequence 4, Appli
16	143.5	4.8	219	2	US-08-557-309B-54 Sequence 54, Appl
17	140.5	4.7	1427	4	US-09-252-991A-27005 Sequence 27005, A
18	138.5	4.6	1061	3	US-09-762-481B-2 Sequence 2, Appli
19	137.5	4.6	442	3	US-08-834-306-52 Sequence 52, Appl
20	137.5	4.6	442	3	US-08-993-674A-52 Sequence 52, Appl
21	137.5	4.6	442	4	US-08-236-976-52 Sequence 32447, A
22	136.5	4.6	700	4	US-09-252-991A-32447 Sequence 6, Appli
23	135.5	4.5	1848	3	US-08-236-791-6 Sequence 6, Appli
24	135.5	4.5	1848	4	US-09-839-996-6 Sequence 6, Appli
25	135.5	4.5	1848	4	US-10-090-505-6 Sequence 6, Appli
26	135.5	4.5	1848	5	PCT-US95-10661A-6 Sequence 2, Appli
27	133.5	4.5	803	3	US-09-083-035-2 Sequence 2, Appli

28	133.5	4.5	1298	2	US-08-590-473-2	Sequence 2, Appli
29	133.5	4.5	1298	3	US-09-259-821A-2	Sequence 2, Appli
30	133.5	4.5	1298	3	US-08-843-659-2	Sequence 2, Appli
31	132.5	4.4	657	3	US-08-893-852A-3	Sequence 3, Appli
32	132.5	4.4	657	3	US-08-821-818-3	Sequence 3, Appli
33	132.5	4.4	657	4	US-09-052-753B-3	Sequence 3, Appli
34	130	4.3	1964	2	US-08-790-912-2	Sequence 2, Appli
35	130	4.3	2052	2	US-08-790-912-3	Sequence 2, Appli
36	128	4.3	309	2	US-08-405-175A-8	Sequence 8, Appli
37	127.5	4.3	532	1	US-08-285-440-5	Sequence 5, Appli
38	127.5	4.3	532	1	US-08-630-349-5	Sequence 5, Appli
39	127.5	4.3	1360	4	US-09-788-657-22	Sequence 22, Appli
40	127	4.2	1002	4	US-09-252-991A-27980	Sequence 27980, A
41	127	4.2	1269	4	US-09-645-456A-15	Sequence 15, Appl
42	127	4.2	1269	4	US-09-435-324A-15	Sequence 15, Appl
43	127	4.2	1269	4	US-09-645-791-15	Sequence 15, Appl
44	127	4.2	1298	4	US-09-645-456A-14	Sequence 14, Appl
45	127	4.2	1298	4	US-09-425-324A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-051-019-2
; Sequence 2, Application US/09051019
; Patent No. 6103229
; GENERAL INFORMATION:
; APPLICANT: KAHMANN, Regine and QUADBECK-SIEGER, Claudia
; TITLE OF INVENTION: Regulatory gene from Ustilago maydis
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, Pentium processor
; OPERATING SYSTEM: Windows 98
; SOFTWARE: WordPerfect version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,019
; FILING DATE: 31-MAR-1998
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-051-019-2

Query Match	5.8%	Score 174.5;	DB 3;	Length 2289;
Best Local Similarity	20.1%	Pred. No. 5.2e-06;		
Matches	120;	Conservative 85;	Mismatches 184;	Indels 209; Gaps 28;
QY	28	GFRTTIAKREGAGDADPFLPPPPQOGLSLRSGRQPKTEVDFLTARRGR	87	
Db	1387	GSRLTDAQR-----DRELD-----KLVERVEDATDPADONKP	1420	
QY	88	SNPVLSDEGPTSCPATDAETASGVS-----SASTRSGPOS-----AS	129	
Db	1421	NANCTCRSMPIAIPSSGAECSCRCVYHLSCKIVRSEVSRAGGWCPCPWGSA	1480	
QY	130	TAVKERPASS-----ERVKGGDDH---DDTSDSDSDGTLTKELQNLRRKREQEPTERPLKG	183	
Db	1481	PFLKKEKAIADIADLSKLVYDQDHRDQFKFLPLEWDAISEVV--AKAKRFETAARMIKT	1538	
QY	184	IQRLRRKREEE-----GPAS-----TVGSEASDT-----	208	
Db	1539	L-SIARRDQKVILAWLRRISGCPVDVTLGPEKVNMLDISENLLALGSGQGAAPV	1597	

QY 209 --VSGVLPKOB-----PENDQGVVSOAGKDDRESKLE-GKAAQDIK 247
 DB 1598 ERKASTPARSDERTETTPLRSSRVPAPADRDG--SPAVDDRRKAKRKGKAKLVF 1655
 QY 248 DERPGDLGRPKPECEGVDNALYICICQPHNNFMICCDCEWFGDCVGVSEARGLL 307
 DB 1656 QEEIG-IGAYR-----DRQPIVCLCHEPESGR-MTACDKCMLWFHNCVRLDDPP--- 1703
 QY 308 ERNGEDVICNCTILOVDETHSTADQ--EAKWRPGDADGTD-----CYSIGTIHQ 358
 DB 1704 NLGNRPICMCCCI-----KAEKRYFOAEVRVFDIGVDPDMLDIRATLSLEK 1753
 QY 359 KSSRDQGIKRIEKAAPNSGKK--KLKIFQPIVTEAPGASKICGPGCHVQAPDSVYCSN 415
 DB 1754 PVS-----KIQSWTSPENKRVILHLEKETPAIHA----- 1782
 QY 416 DCILKHAATMKFLUSSKOKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 475
 DB 1783 -----EVSQITKARLESPTPSKARVSLGRSDSIST---PA---K 1818
 QY 476 ETTVKAVVVPARSEALG-----KEAACSSSTPSWASDENYNAVXP---EKTAAPSP 524
 DB 1819 ESGAVPYAAPVPSEAVRGIVPALTPAADSPASRGNDSDFAAASPPLWDAKTGESP 1876
 RESULT 2
 US-09-252-991A-18387
 ; Sequence 18387, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 18387
 ; LENGTH: 1075
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-18387
 Query Match 5.5%; Score 164.5; DB 4; Length 1075;
 Best Local Similarity 21.5%; Pred. No. 1.4e-05;
 Matches 126; Conservative 76; Mismatches 228; Indels 155; Gaps 26;
 QY 7 PSNEEAPKAKPTSKFRKTWGFRTTIAKREGAGDAEADPLE--PPPPQQQLGLSLERS 64
 DB 553 PQTEAPVPEAKPMPSPSLFQGLVKSIVGLFACKQPAKPARTSKPAERCTQDEREN 612
 QY 65 GRQPKRTERVEQFTIARRGRSMFVSLSDSEPTSCPATDAHTASGVSASHTSG 124
 DB 613 GQONRRDGRD---GNRRDERKPREEAERQPRE-----ERAERNREERSERRK 662
 QY 125 PQGASTAVKRPASSERKVGSDHDDTSDSDGLTKBLQNLRLKREQPTER---P 180
 DB 663 ERA-----ERPAERER-----QPREGREERAEHTPREERQ 693
 QY 181 LKGIQSR-LKKREEREPARTVSEASDVTVEGVLPSKQEPENDQGVVSOAGKDDRESK-- 237
 DB 694 REGSGREERSERRERARERARERARERARERARERARERARERARERARERARER 749
 QY 238 --LEGKA-----AQIDKDBEPDGLGRPKPECEGYDPNLYCICQPHNNFMICCDRC 288
 DB 750 AALEABALPNDESLEQDDQDDTGE--RPERRSRQ-----RRRSNR--ERQ 793
 QY 289 EWFHGDVGVSEARGKLLERNSDICYNCTI-----LQVQDE-----THS 330

DB 794 RE-VSGVEG-SEAT-----DNAAPLNTVAAAAAGIIVASEAVEANVTQAPATTS 843
 QY 331 ETADQOQAKRPGDADGTDCTISGTTIEQKSSSQGKGR-----IEKAANPSGKK 381
 DB 844 EAASETTA-----SDTASTSEAVTQDADSEANTGETADIEAPVTVSVWRDEADQST 897
 QY 382 LKIQPIVTEAPGASKCI-----GPGCHVQAPDSVYCSNDCILK 420
 DB 898 LLVQAQATEEAPFASGVSRESREDAESAVOPATEAAEVAAPVVEVAAPSFAATEPTPA 957
 QY 421 HAA-----ATMKFLLSGGEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 476
 DB 958 IAAVAPANATGALNDPREKRIQREAEERLAREAAAAAQAQA---PAVEEIPAVASEE 1014
 QY 477 TTVKAVVVPARSEALG-----KEAACSSSTPSWASDENYNAVXP----- 521
 DB 1015 ASAQEE---PAAPQA---EBITQADVPSQA-DEAQAQVQAEPEAS 1052
 RESULT 3
 US-08-769-309A-5
 ; Sequence 5, Application US/08769309A
 ; Patent No. 5741890
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klauck, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/769,309A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5741890and, Greta E.
 ; REGISTRATION NUMBER: 35,302
 ; REFERENCE/DOCKET NUMBER: 27866/33451
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-474-6300
 ; TELEFAX: 312-474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1780 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-769-309A-5
 Query Match 5.4%; Score 161.5; DB 1; Length 1780;
 Best Local Similarity 19.4%; Pred. No. 5.1e-05;
 Matches 133; Conservative 76; Mismatches 221; Indels 255; Gaps 28;
 QY 3 DKGDPSNEEAP--KALKPTS-----KEFRKTW-GFRRTT----- 33
 DB 266 EREGKEQKEKPEKPSABSPTSPVTSETGTFKFKFTQGNAGWRKKTFRKPKEDVEASEK 325
 QY 34 -----TAKREGAGDAEADPLEPP-----PQOQLG 58
 DB 326 KKEQPEKVTERTDGRKAEVASEKLTASEQAHQFQPAESAHEPRLSAEYKVELPSEQV- 384


```

; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-257-179-80

Query Match      5.4%; Score 161; DB 4; Length 238;
Best Local Similarity 33.7%; Pred. No. 3.3e-06;
Matches 33; Conservative 17; Mismatches 18; Indels 30; Gaps 4;

QY 230 GNDRESKL-----EGKAAQIDKEEPDGLGRPKPECEGYDPNALYICRQPHN-NRF 281
Db 32 GITEKAKMDVYICNDCEKAGGSSE-----LYCICETPYDEQF 73
QY 282 MTCDCREWFHGDGCVGISEARGRLIRNGEYICPNC 319
Db 74 YIGCDRCQWYRGRCVGLQSAELT-----DEYVCPQC 107

RESULT 6
US-10-164-595-56
; Sequence 56, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 56
; LENGTH: 779
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-56

Query Match      5.3%; Score 158.5; DB 4; Length 779;
Best Local Similarity 21.4%; Pred. No. 2.9e-05;
Matches 104; Conservative 58; Mismatches 191; Indels 134; Gaps 19;

QY 20 SKFRTWGFRTTIARAGAGADADPLEPPPPQQLGLSLRSGRQPKRTVERQPLT 79
Db 278 SREISK---FRDTHKKLEKGGKKE-----RQIEKRRERERE 315
QY 80 IARRRRRSMFVSLDSGSEPTSCPATDAETASGVSASSETRSGPQSASTAVKERPASS 139
Db 316 RERERER-----ERREERERERERERERERERERERERERER 357
QY 140 EKVGGDDHDDTSDSDGLTLKELONLRKRQEPTERPLKIGISRLRK-----191
Db 358 ERERERKRRDEDEEDAYEREKRLERLKB-BAAYQER-LKNWIRERKKTREYKKA 415
QY 192 RREEGPAETVGEA---SDTVGEVLPKQBPENDQGVSGAGKDDRESKLEGKAAQDIKD 248
Db 416 EREERERREKAKAKLKEFLDYDDDDPKYRGSAQKRLRDRKEMEA-----D 468
QY 249 EEPGLGRPKPECEGYDPNALYICRQPHNRFMI CDRCEWFHGDGCVGISEARGELLE 308
Db 469 ER--DRKREKEELE-----EIRQLLIA 488

; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-257-179-80

Query Match      5.4%; Score 161; DB 4; Length 238;
Best Local Similarity 33.7%; Pred. No. 3.3e-06;
Matches 33; Conservative 17; Mismatches 18; Indels 30; Gaps 4;

QY 230 GNDRESKL-----EGKAAQIDKEEPDGLGRPKPECEGYDPNALYICRQPHN-NRF 281
Db 32 GITEKAKMDVYICNDCEKAGGSSE-----LYCICETPYDEQF 73
QY 282 MTCDCREWFHGDGCVGISEARGRLIRNGEYICPNC 319
Db 74 YIGCDRCQWYRGRCVGLQSAELT-----DEYVCPQC 107

RESULT 7
US-08-801-308-1
; Sequence 1, Application US/08801308
; Patent No. 6368790
; GENERAL INFORMATION:
; APPLICANT: Scott, Robert E.
; TITLE OF INVENTION: cDNA ENCODING P2P PROTEINS AND USE OF
; TITLE OF INVENTION: P2P cDNA-DERIVED ANTIBODIES AND ANTISENSE REAGENTS IN
; TITLE OF INVENTION: DETERMINING THE PROLIFERATIVE POTENTIAL OF NORMAL,
; TITLE OF INVENTION: ABNORMAL AND CANCER CELLS IN ANIMALS AND HUMANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates, P.C.
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,308
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.6435P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-308-1

Query Match      5.3%; Score 157; DB 4; Length 1404;
Best Local Similarity 20.5%; Pred. No. 9.1e-05;
Matches 133; Conservative 77; Mismatches 228; Indels 212; Gaps 30;

QY 8 SNEERAPK--ALKPTSKFKTKWGFRTTIARAGAGADADPLEPPPPQQLGLSLRSG 65
Db 851 AKKEATKIDSVKPPSSSQKD-----EKVTGTPKAKHSKAKDTRRSQPR-----TRRSK 900
QY 66 R-----OPKRTVERVQPLTI-----ARRGRSRMFVSLDSG-----RPT 100
Db 901 RTVPKTSQKQSVKTRPRSLKINYLAREKREKREKRSKVDKDFSSSMKIKVGT 960
QY 101 SCPATDAETASGVSASSETRSGPQSASTAVKERPASSKVKGGDDHDDTSDSDGLT 160
```

Db 961 EIVKPSKXVGDVEXLETT---PEKDIASSSTPA---KKIKLNR-----T 1003
Qy 161 LKELONLRKRQEPTEPLKGIQSL-----RKREEGPAETV----- 201
Db 1004 GKIKGNAENASTTKPSEK-LESTSSIKIKQKVKGAKRVAGSESSSTLVYTTST 1062
Qy 202 -----GSEASDVTVEGLPSKOPEND-----QGVVSOAGKDDRESKLEGA 242
Db 1063 GGSFVRKSEBKTDTKRTVINTMEBYNNDNTAPAEVDLIIMIQVPSKWDKDFSEEB--- 1119
Qy 243 AQDIKDEP-GDLGRP-----KP-----BCEGYDPNLYCICRQPHNRFWICC 285
Db 1120 -DVYKTPQIQSVGKPSIIFKNVTKESATAKYTEKESQPEKLOKLPKEASH----- 1170
Qy 286 DRCEWPHGDCVIGSEARGLLER--NGEDYICPN-----CTILQVODETHSET 332
Db 1171 ELWQHELRSKGSNSSEKRAKREHSGSEKNDPKKSGNQPKBSTVDRLSOGHFKT 1230
Qy 333 ADQOEAQRFGDAGTDTCTSIGTIBQKSDGIGIKRIEKAANPSGKKUKIFQPVIEAP 392
Db 1231 LSGSSKETR-----TSEKHSVRGSSNKDFTFGRDKVD----- 1264
Qy 393 GASKCIGPGCHVAQPSVYCSNDCILKHAATMKFLSSGKQKPKPB-----KMKMKP 447
Db 1265 -----YSDRYSSKRRDERGELARRDSPPRKESLSQKSKIRE 1305
Qy 448 EKPSLPKCGAAGIKISSV--HKRPAPEKETTAKVAVVVPARSHALGKAACESSTPSW 505
Db 1306 ER-DLPKKGAKSKNSPPRDKKPHKAPYETKR-----PCSETKP-- 1347
Qy 506 ASDHYNNAVPEKTAAPS-----BSLLYKVMVHLGVGLDPSRFWIAIPW 551
Db 1348 -VDKN-SGKERKHAERNGKSSGANCHVYILTRQTL-PWRSWLLGRW 1394

RESULT 8

US-10-164-595-80
; Sequence 80, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-80

Query Match 5.1%; Score 152.5; DB 4; Length 735;
Best Local Similarity 20.8%; Pred. No. 9e-05;
Matches 108; Conservative 75; Mismatches 227; Indels 109; Gaps 20;
Qy 2 DDKGDSNEAPKAIKPTSKFRTWGR---RTTIK---REGAGDAEADPLEPPPPQ 54
Db 163 EDKRLISREISK-FRTHKKLEEEKKKEKEQIEKERRERERERERERERERERER 221
Qy 55 QQLGLSLRRSGRQPKRTERVEQFLTIARRGR-----RSMFVSLSDSGPTSCPATDA 107
Db 222 EREREREKERERERERDRDRTKDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR 281
Qy 108 ETASEGSVSASSTRSGPOSASTAVKRPASSEKVKGGDDHDDTSDSDGLTKELQNR 167
Db 282 DRER 341
Qy 168 LRKREQEPTEPLKGIQSLRK---RREGPAETVGEA---SDTVEGVLPSPK 216
Db 342 LREK-EAAYQER-LKNWEIRKKTREYKEAERERERERERERERERERERERERER 399

Qy 217 QEPENDQGVVSOAGKDDRESKLEGAQDIKOBEPGLGRPKPCEGYDPNLYCICRQP 276
Db 400 DDPKYTRGSALQKLLADREKMEA-----DER--DRKREKEELE----- 436
Qy 277 HNRFMICCDRCBETHGDCVIGSEARGLLERNGEDYICPNCTILQVODETHSETADQ 336
Db 437 -----BIRQLLAEGHPD---PDAELQRMQB---AERRRQP 467
Qy 337 EAKWRPGDADGTCTSIGTIBQKSSDQIKRIEKAANPSGKKLK-IPQVIEAPGAS 395
Db 468 QIKQEPSESEER-----EKQEKERERERERERERERERERERERERERERER 519
Qy 396 KCIPOGCHVAQPSVYCSNDC--ILKHAATMKFLSSGKQKPB---KPKKKKKMK--P 447
Db 520 SASGN-----ATENTPDESFCGIIIPHE-----NSPDQOQPEHRPKIGLSLKIGAS 567
Qy 448 EKPSLPKCGAAGIKISSVHKRPAPKPKETTAKVAVVVP 486
Db 568 NSPQGPNSVKRKLVDVSNFNFEDSDSDVPRKRKLVP 606

RESULT 9

US-10-164-595-79
; Sequence 79, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 79
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-79

Query Match 5.1%; Score 152.5; DB 4; Length 784;
Best Local Similarity 20.8%; Pred. No. 9e-05;
Matches 108; Conservative 75; Mismatches 227; Indels 109; Gaps 20;
Qy 2 DDKGDSNEAPKAIKPTSKFRTWGR---RTTIK---REGAGDAEADPLEPPPPQ 54
Db 212 EDKRLISREISK-FRTHKKLEEEKKKEKEQIEKERRERERERERERERERERER 270
Qy 55 QQLGLSLRRSGRQPKRTERVEQFLTIARRGR-----RSMFVSLSDSGPTSCPATDA 107
Db 271 EREREREKERERERERDRDRTKDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR 330
Qy 108 ETASEGSVSASSTRSGPOSASTAVKRPASSEKVKGGDDHDDTSDSDGLTKELQNR 167
Db 331 DRER 390
Qy 168 LRKREQEPTEPLKGIQSLRK---RREGPAETVGEA---SDTVEGVLPSPK 216
Db 391 LREK-EAAYQER-LKNWEIRKKTREYKEAERERERERERERERERERERERERER 448
Qy 217 QEPENDQGVVSOAGKDDRESKLEGAQDIKOBEPGLGRPKPCEGYDPNLYCICRQP 276
Db 449 DDPKYTRGSALQKLLADREKMEA-----DER--DRKREKEELE----- 485
Qy 277 HNRFMICCDRCBETHGDCVIGSEARGLLERNGEDYICPNCTILQVODETHSETADQ 336
Db 486 -----BIRQLLAEGHPD---PDAELQRMQB---AERRRQP 516
Qy 337 EAKWRPGDADGTCTSIGTIBQKSSDQIKRIEKAANPSGKKLK-IPQVIEAPGAS 395
Db 517 QIKQEPSESEER-----EKQEKERERERERERERERERERERERERERERER 568
Qy 396 KCIPOGCHVAQPSVYCSNDC--ILKHAATMKFLSSGKQKPB---KPKKKKKMK--P 447

Db 569 SASGN-----ATPNTPGDSCGIIIPHE-----NSPDQQPPEHRPKTGLSLKLGAS 616
Qy 448 EKPSLPKCAQAGIKISSVHRKPAPEKKEKETTIVKXAVVP 486
Db 617 NSPGQPSVVKRKLKLPVDSVFNKPFEDSDDDVPRKRLVP 655
RESULT 10
US-10-164-595-54
; Sequence 54, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 54
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-54
Query Match 5.1%; Score 152.5; DB 4; Length 843;
Best Local Similarity 20.8%; Pred. No. 0.00011;
Matches 108; Conservative 75; Mismatches 22; Indels 109; Gaps 20;
Qy 2 DKGDPNNEAPKAIKTSKFRKTWFR-----RTTIAK-----REGAGDAEADPIPPPPQ 54
Db 271 EDKDLISREISK-PRDTHKLEBKKGKKEKQBIKERERERERERERERERERERERERER 329
Qy 55 QQLGLSLRSRGPQKTERVQFLTIARRGR-----RSMVSLSDSGEPTSCPATDA 107
Db 330 ER 389
Qy 108 ETASGVSSESATSGSQSASTAVKRPASSEKVGKGGDDHDDTSDSDGLTLKEQNR 167
Db 330 DRER 449
Qy 168 LRRKEQEPTEPLKIGOSRLKK-----RREGPAETVGSBA-----SDTVEGLPSK 216
Db 450 LREK-EAAYQER-LQNWIREKKTREYKEAEERERERERERERERERERERERERERERER 507
Qy 217 QEPENDQGVSAQKDDRESKLEKGAQDIKDEEPGLGRPKPECEGYDPNALYICICRQP 276
Db 508 DDPKYRGSALQKRLDRKEKEA-----DER--DRKREKHELE-----544
Qy 277 HNNRPMICCDRCBWFHGDVCGISARGRLLRNEDYICNCTILOVDTHSETADQQ 336
Db 545 -----BIRQLLAEGHPD-----PDASLQRMQE-BAERRRQP 575
Qy 337 RAKWRPGADGTDCTSIGTIOKSSSEDGIGRIEKAANPSKKLKL-IPQVTEAPGAS 395
Db 576 QIKQPESEHEE-----EKQKEKREEMEEBEPQKPCLKTLPIASAPVS 627
Qy 386 KQIGPGCCHVAPDSVYCSNDC--ILKHAAATFMKLSGKQKQP---KPKRKKMK---P 447
Db 628 SASGN-----ATPNTPGDSCGIIIPHE-----NSPDQQPPEHRPKTGLSLKLGAS 675
Qy 448 EKPSLPKCAQAGIKISSVHRKPAPEKKEKETTIVKXAVVP 486
Db 676 NSPGQPSVVKRKLKLPVDSVFNKPFEDSDDDVPRKRLVP 714

RESULT 11

US-08-978-277A-4

; Sequence 4, Application US/08978277A

; Patent No. 6582956

; GENERAL INFORMATION:

; APPLICANT: Gelman, Irwin H.

; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/978,277A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/665,401
; FILING DATE: 18-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S
; REGISTRATION NUMBER: 26,154
; REFERENCE/DOCKET NUMBER: A30558 - 165/34008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2558
; TELEFAX: 212-765-2519
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1596 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-978-277A-4

Query Match 5.1%; Score 152.5; DB 4; Length 1596;

Best Local Similarity 19.6%; Pred. No. 0.00027;

Matches 141; Conservative 78; Mismatches 236; Indels 263; Gaps 32;

Qy 2 DKGDPNNEAPKAIKTSKFRKTW-GFRRTTIAK-----EGAGDA 43
Db 270 ESPSPVNSSTTSFK---KFFTHGWAGNKKTSFKSKEDDLETAEKKEQAEKVDDE 326
Qy 44 BADPLEPP-----PPQOLGLSLRSRGP---PKETERVEQ 76
Db 327 EKEKTEPASEQSPAEPTDQARLSADYKVELPLDQVG-DLEASSEKCAPLATFVDE 385
Qy 77 FLTIARRGRRRSPVSLSDSGEPTSCPATDAE--TASEGSVES-----ASSTRSGPOSAS 129
Db 386 KME-AHQEVVAHVHSIVKTEBEEQGGGAEQGVVVEGTGESLPPEKLAEPQVPOEAE 444
Qy 130 TAVKERPASSEKVKGGDDHDDTSDSDGLTLKE-----LQNRLEKKEQEPTE 178
Db 445 PA-BELMKSRMCMVSGGDHTQLTDLSPERKTLPKHPGIVSEVEMLSQERIKVQGSPLK 503
Qy 179 R-----PLKGIQSRLEKRR---EEGPAB-----TVGSASD-----207
Db 504 KLFSSSGLKLSGKKQKKGKGGGDEEPGEVQHHTSPESADQKGSASSPEEPET 563
Qy 208 --TVEGVLPKQEPENDQGVVSQAK-----DRESKLE 239
Db 564 TLEKGPLEAPQDGEAEGETTSDGEKREKRGITPWASFKKMVTPKKVRPSESDKEELE 623
Qy 240 -----GKAAQDIKDEEPGLRKPKECEGYDPNALYICICRQPHNRFMICCRCE 289
Db 624 KVKSATLSSTSTVSEMQDEKVTGVEEQKE-EPKRR-----VDTSV 664

Qy	290	EWFHGDVCVIGSEARGRLRNGEDVICPNTCTILOVDETHSET--ADQEAKWRFQDAD-	346
Dd	665	SWEALICVGSSKGRARKASS-----DDEGGPTLGGDSHREASAKDKEA	710
Qy	347	GTDCTSIGTIOKQSSDOGIEKRIEKAANPSGKKLKIFQPVIRAPGASKICPGCCCHA	406
Dd	711	GTDVAVPASTQDQAQGSSSP--EPAGSPS-----EGEGVS-----	744
Qy	407	QPDSVSYNCDCILKEAAATMKFLSSGKQKPKEKMMKPKPSLPKGCAQGIKISV	466
Dd	745	-----TWSPFKLVTPRKKSCLKLEKAEDSSVBQL-----ST	777
Qy	467	HKRPAPEKETT VKKAVVPVPARSEALGK--EACCESSSTPSWASDNH-----YNA	513
Dd	778	BIESPREESWSYSIKFIPERRERKRDGQRQATVEDSGPVEINDDPNVPVAVPISEYNA	837
Qy	514	VKPKRTAPS-----PSL-----LYKMVH-LGVGLLDPSRSP-----WIA	548
Dd	838	VEREKWEAQGNTELPLLGAVVVSBELSKTLVHTVSVAVIDGTTRANTSVIERSPPSWIS	895

```

RESULT 12
US-09-976-594-726
? Sequence 726, Application US/09976594
? Patent No. 6673549
? GENERAL INFORMATION:
? APPLICANT: Furness, Michael
? APPLICANT: Buchbinder, Jenny
? TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
? FILE REFERENCE: PA-0041 US
? CURRENT APPLICATION NUMBER: US/09/976,594
? CURRENT FILING DATE: 2001-10-12
? PRIOR APPLICATION NUMBER: 60/240,409
? PRIOR FILING DATE: 2000-10-12
? NUMBER OF SEQ ID NOS: 1143
? SOFTWARE: PERL Program
? SEQ ID NO 726
? LENGTH: 2468
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? OTHER INFORMATION: Incyte ID NO. 6673549 4295277CD1
US-09-976-594-726

```

Query Match	5.0%;	Score 149.5;	DB 4;	Length 2468;
Best Local Similarity	20.0%;	Pred. No. 0.00091;		
Matches 115;	Conservative 88;	Mismatches 190;	Indels 183;	Gaps 27;

Qy	2	DDXGDPNVEAPKAIKPTSEKFTWGPFRRTIAKREGAGDAEADPLEPPPPQOQLGLSL	61
Db	691	EEKKSP--KGEVKETTP--KEVKKE-----VKKEKGVKKEKEPKKEIKKLPKDA	739
Qy	62	RRSG-----RQPKRTERVEQ-----FLTAIRRRGRSPVSLIE--	94
Db	740	KKSSTPLSEAKPAALPKVPKKEESVKQDSVAAKPKKGIKVIKKEGKAEEAAVAAV	799
Qy	95	DSGEPTSCPATDAETASGSEVESASETRSGQOSASTAVKRPASEKVKGGDDHDDTSDS	154
Db	800	GTGATTAAVMAAGTAAIGPAKEALEASRLSSPEDITKD-----FEEL-AEEDVVTXDI	854
Qy	155	DSGLTLKELQNLRRKREQRTERPLKGIOSRLRKRRB--EGPAETVGSSEASDTVEGV	212
Db	855	KPQ--LELIDDEKLI---KETBP-----VBAVYQKEREVTYKGPAS--PDEGITTTEGE	902
Qy	213	LPSKQEPENDGVVSQAGKDRESKLEGKAAQDIKDPEPGLGRKPECEGYDPNALYCI	272
Db	903	GECEQTPEELEPEVKQ--GVDDIE--KFDEGAGFPSESSSTGDYBEKAEETAEAEPP--	954
Qy	273	CROPHNNRFJCCDRCEWFHGDGVGISIEARGRLLENGEYDIPCNCTILOQVDETHSET	332
Db	955	-----EDGESEHVC-----VSASKEHSPT	972

```

333 ADOQBAKWRPGDAG--TDCTSIGIIBOKSSD-----QGIKGRIEKKAANPSGKKK-- 381
973 EDESASAK---AEADAYIRKRESVAGSDRAEEDMDAIEKGAEOSEBEADBEDKAEDA 1029
382 -----LKIPQPVIEAPGASKCIG-----PGCCHVAQPDVY 412
1030 RBEYEPKMEAEADYMAVVDKAAEAGGAEQYGLTPTTKQLGAQSPG----- 1078
413 CSNDILKHAATW--KPLSSSGKEQK--PPPKKMMKKPKPSLPCGAQAGIKISSVHK 468
1079 -----REFASSIHDTLLFGGSESEATASDENREDQPEFTATSGYTOSTBISS--- 1128
469 RPAPEKETTIVKAAVVVPARSEALGKEAACESSTPS 504
1129 EPTFMDMSTPRDVMASDETNE-----ETESPS 1156

RESULT 13
US-08-714-741-32
Sequence 32, Application US/08714741
Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yother, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooke-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Prommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 8991 amino acids
Type: amino acid
Strandedness: single
Topology: linear
MOLECULE TYPE: amino acid
US-08-714-741-32

```

Query Match 5.0%; Score 149.5; DB 4; Length 8991;
Best Local Similarity 20.5%; Pred. No. 0.0056;
Matches 129; Conservative 73; Mismatches 260; Indels 167; Gaps 25;

Db 7246 TQOPPKAEKAPAPAKTGNKQENGMWYFYNTDGSNGEQAYRAAEGDLAAKQABLEKT 7305
Qy 58 GLSRRSGROPKRTVERVOFTIARRGRRRMFVLSDSGHTSCPA----- 104
Db 7306 EADLKAVNEPKAPAPAE-----TPAPEAPQPKAPAPAPAPAPAPKPKPA 7353
Qy 105 -----TDAATASEGSVESASE-----TRSGPOSASTAV-----KRPASSEKVKGGDDH 148
Db 7354 EOPKAEKTDQOAEEDYARRSEHEYNRLTQOPPKAEKAPAPKPEQAPAPAPKSKGBOA 7413
Qy 149 DDTSDSDSGGTLKELQRLRRKROEPTERPLKIGISRLKRRREGPA-----ETVGESEA 205
Db 7414 EYRSAAGDLAAKQV-----ELEKTEADLK-----KAVNEPKAPAPETPAPEA 7459
Qy 206 SDTVEGVLPKQ-----EPENDQGVVSGAKND-DRESKLEK-----AAQDLKDE 249
Db 7460 PAEQPKAPAPQAPAPAPKPEKAPAEQPKAEKADQOAEEDYRRSEHEYNRLTQOPPKAE 7519
Qy 250 BPGDLGRPKPECEGYDPNLYCICQPHNNRPMICCDCEWPFHGDVGVISEARGRLIER 309
Db 7520 KPAP--AQPEQAPAPAKSLKDESDSE-----DVKEGFRAPLQSELDQAQKLSK 7570
Qy 310 NGE--DYICPNCTILQVODETHSETA-----DOEAKRPGDAGDCTDSIGTIEQ-----K 359
Db 7571 LEELSDKI-----DELDAETAKLEKQVEDFKXSDGBOAG-----QYLAABEDLIAK 7617
Qy 360 SSEDQGIKGRIEKAANPSGKKLIP--QPVTAPAGASKCIGPGCHVA----- 406
Db 7618 KAELEQTADLKAVNEPKAPAPAPETPAPEAPAEQPKAPAPETPAPEKPEKPAEQK 7677
Qy 407 --QPSVVCNSNDCLLKHAAATWKFLSSGKEQPKPKPKMKM--KPEKPSLFCGQAQAIK 462
Db 7678 PKPADQOAEED-----YARRSEHEYNRLTQOPAPAPAKPEQAPKPEKTAPEPTQEKDAE 7733
Qy 463 ISSVH-----KRAPEKCTTVKAVVVPARSEALGKE 495
Db 7734 IAKLEKQVYFKTDAEQTEQVLAARQDLADKKALEKTEADLKAVNEP-----EKPAEE 7790
Qy 496 ACESSTPSWASHDHNNAVPEKTAAPSP 524
Db 7791 TPAPAPKPEQAPAEQPKAPAPAPAPAPAPK 7819

RESULT 14
US-07-853-913-2
Sequence 2, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
APPLICANT: Lendahl, Urban
TITLE OF INVENTION: Nestin Expression As An Indicator of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA: US 07/201,762
APPLICATION NUMBER: 02-JUN-1988
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-913-2
Query Match 5.0%; Score 148.5; DB 1; Length 1805;
Best Local Similarity 19.8%; Pred. No. 0.00072;
Matches 132; Conservative 105; Mismatches 223; Indels 207; Gaps 36;
Qy 10 EAPKATKPTSKPRKTWG---FRRTIYAKREGAG-----DAADPLEPPPPQOOLG 58
Db 521 EETQDSQPLQKXETLKALGEEPLMSLKIONYETAGKENCNSTEHLGTLGGPKKKQIP 580
Qy 59 L-SLRSS-----GROPKRTE-----RVEQFLTAIR----- 82
Db 581 LKSLKEKNVSEKTLGVVPLVSELLOKEDTRIEDQELMSPKGTLLKRFSSLGKESQEVVR 640
Qy 83 --RRGR-RSMFVSLSDSEPTSCPATDAETASGVSVSASSETRSGPOSASTAVKE--RPA 137
Db 641 PSKEGNLESMTAFKESQHPPLGPGAE--DOMLEPLVB--KEDQSPFRSPDEEDQACRPL 697
Qy 138 SSEKVGKGDHDDTSDSDGLTKELQNLRRRREOEPTERPLKIQSLRKKRRREGP 197
Db 698 QKE-----KQELPYERAEQ-----QILRLIETKESQESLESPEEDQEAGSLSQKEN-- 745
Qy 198 AETVG--SEASDTVBGVLPKQ-----EPENDQGVVSGAKNDRES-----KL 238
Db 746 QEPLGYEAEQMLERLITEKESQESLSKSPENQRIKGPLERENOKSLYLEENQFTFVPL 805
Qy 239 EGKAAQDIK---DEEFGDLGRPKPECEGYDPNLYCICR---OPHNNRPMICCDRCCEW 291
Db 806 ESRNQRPLRSLEVEEERQIVKP---LEKVSQDSLSGLAEENVQP--LRYL-----EE- 853
Qy 292 FHGDCVGIS-----EARGELLRNGEDYICPNCTILQV-----ODE---THSETAD 334
Db 854 --DDCINKSLLEDKTKSLGSLIEDRNGDSIIIPQSETOVSLRPEEEDQRIVNHLEKES 911
Qy 335 Q-----QBAKWRPGDADGTCISIGTIEKSSSEDQGIKGRIEKAANPSGKKLQIFOP 387
Db 912 QFSSRSSEEEQVMSRLEGENHESLSVEK---EDQVVSQLEKESQDSGKSL----- 962
Qy 388 VIEAPGASKCIGPGCHVAQPDVSVYCSNDCILKHAATMKTLLSSCKEQKPKPKERKYM-- 445
Db 963 ---EDESQETFGP-----LEKNAESLSLAGQDEQKLEQSTQOTL 1002
Qy 446 -----KPEK--PSLPK-----CQAQAGIKLSSVHKPAPEKKTETVKK 481
Db 1003 RAVGNEQMAVSPPEKVDPELPKPLGNDQEIARSLEKQESLSVSL-----KKGLETYKS 1057
Qy 482 ---AVVVPARS--EALGKEACACESSTPSWASHDHNNAVPEKTAAPSPSLLYKCMYHLGV 536
Db 1058 LETEIIIBLETABEDLERKSIDTOEPLWSTVEARETVPEPDEPPG-----SL 1106
Qy 537 GLLDPSR 543
Db 1107 GSVDENR 1113

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:51:17 ; Search time 45 Seconds
(without alignments)
3940.472 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDKGDPSNEAPKAIKPTS.....RSFWIAIPWACGLGVAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329.5	11.0	2016	5 Q9VG78	Q9VG78 drosophila
2	217	7.3	663	5 Q9W352	Q9W352 drosophila
3	202.5	6.8	366	5 Q8T3Y1	Q8T3Y1 drosophila
4	197	6.6	2649	5 Q9W0T2	Q9W0T2 drosophila
5	197	6.6	2669	5 Q9SVB8	Q9SVB8 drosophila
6	197	6.6	2669	5 Q9W0T1	Q9W0T1 drosophila
7	194	6.5	3146	5 Q9VUB5	Q9VUB5 drosophila
8	191	6.4	2764	4 Q7Z7D6	Q7Z7D6 homo sapien
9	187.5	6.3	542	5 Q918F0	Q918F0 euplotes cr
10	187	6.3	443	11 Q8C969	Q8C969 mus musculus
11	187	6.3	473	11 Q8C9E0	Q8C9E0 mus musculus
12	187	6.3	17352	5 Q9SYM2	Q9SYM2 procambarus
13	182	6.1	563	13 Q7SZX6	Q7SZX6 brachydanio
14	182	6.1	563	13 Q7SZX4	Q7SZX4 brachydanio
15	181	6.1	513	5 Q9SPU8	Q9SPU8 chironomus
16	181	6.1	720	5 Q7Z2C9	Q7Z2C9 drosophila

17	178	6.0	606	6 Q28687	Q28687 oryctolagus
18	178	6.0	2400	4 Q8IWP2	Q8IWP2 homo sapien
19	178	6.0	2416	4 Q8IWP1	Q8IWP1 homo sapien
20	178	6.0	2432	4 Q8IWP0	Q8IWP0 homo sapien
21	178	6.0	2448	4 Q8IWN9	Q8IWN9 homo sapien
22	178	6.0	2464	4 Q8IWN8	Q8IWN8 homo sapien
23	178	6.0	2480	4 Q8IWN7	Q8IWN7 homo sapien
24	177	5.9	2464	4 Q86SQ1	Q86sq1 homo sapien
25	176.5	5.9	1417	3 Q871Y7	Q871y7 neurospora
26	175	5.9	424	3 Q74508	Q74508 schizosacch
27	175	5.9	545	11 Q8VDN7	Q8vdn7 mus musculus
28	175	5.9	5327	5 Q76891	Q76891 drosophila
29	174.5	5.8	669	11 Q35540	Q35540 mus musculus
30	174.5	5.8	687	11 Q9X75	Q9x75 rattus norv
31	174.5	5.8	2289	3 Q9HFW4	Q9hfw4 ustilago ma
32	174	5.8	5412	5 Q9W596	Q9w596 drosophila
33	172	5.8	1135	6 Q8HZN3	Q8hzn3 canis famil
34	171.5	5.7	700	4 Q9BUV3	Q9buv3 homo sapien
35	171.5	5.7	1110	13 Q91255	Q91255 petryozon
36	170	5.7	510	5 Q45407	Q45407 caenorhabdi
37	170	5.7	878	13 Q8AVW4	Q8avw4 xenopus lae
38	169.5	5.7	620	10 Q8LNZ4	Q8lnz4 nicotiana t
39	169	5.7	678	11 Q91L92	Q91l92 mus musculus
40	169	5.7	990	13 Q91803	Q91803 xenopus lae
41	169	5.7	1893	11 Q8CJ14	Q8cj14 rattus norv
42	169	5.7	4969	11 Q8CF91	Q8cf91 mus musculus
43	169	5.7	5165	11 Q8CF92	Q8cf92 mus musculus
44	167	5.6	1408	5 Q9W0C9	Q9w0c9 drosophila
45	166.5	5.6	461	11 Q9QX76	Q9qx76 rattus norv

ALIGNMENTS

RESULT 1

Q9VG78					
ID	Q9VG78	PRELIMINARY;	PRT;	2016	AA.
AC	Q9VG78;	(TREMBlrel. 13, Created)			
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)			
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)			
DE	01-OCT-2003	(TREMBlrel. 25, Last annotation update)			
DE	CG6525	protein.			
GN	SPP OR CG6525.				
OS	Drosophila melanogaster (fruit fly).				
OC	Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Berkely;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Branden R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriel J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Beres P.V., Berwan B.P., Shandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Belcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fowler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glick A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.I., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				

STRAIN=Berkeley;
 MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Banton D.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.S., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.C., Dunn P.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Fostel C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson J.A.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 RA [3]
 RA SEQUENCE FROM N.A.
 RA Celinker S.B., Adams J.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrieria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Parasag V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RA "Sequencing of *Drosophila melanogaster* genome.";
 RA Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
 RA [4]
 RA SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berwan B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of *Drosophila melanogaster* genome.";
 RA Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
 RA [5]
 RA SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases.
 RA [6]

```

RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AY089443; AAL90181.1; -.
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Housck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkliv G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacile J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Bazon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreanek D., Farfan D.,
RA Ferreira S., Frise B., Galle R.P., Garg N.S., George R.A.,
RA Gonzalez M., Housck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacile J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.I., Bergman C., Bernan B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AB003467; AAF47361.2; -.
RA FlyBase; FBgn0000541; E(bx).
RA GO; GO:0005634; C:nucleus; IEA.
RA GO; GO:0003677; F:DNA binding; IEA.
RA GO; GO:0005489; F:electron transporter activity; IEA.
RA GO; GO:0006119; F:electron transport; IEA.
RA GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
RA InterPro; IPR000637; AT hook.
RA InterPro; IPR001487; Bromodomain.
RA InterPro; IPR00345; CytC heme_BS.
RA InterPro; IPR004022; DDT dom.
RA InterPro; IPR006209; EGF-like.
RA InterPro; IPR001365; ZnF-PHD.
RA Pfam; PF02178; AT_hook; I.

```

DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2649 AA; 298507 MW; COD7F7E015EA5403 CRC64;

Query Match 6.6%; Score 197; DB 5; Length 2649;
Best Local Similarity 19.0%; Pred. No. 0.0001;
Matches 82; Conservative 75; Mismatches 139; Indels 136; Gaps 15;

Qy 49 EPPPPQOGLSLRRSGRQPKTER-----VEQFLTIARRGRRRMPVSLD----- 95
Db 2197 EHDEPTNLGLDISETDLNKKQNESFVVTGRYIQKSISNALKQ--NLSPELEKLVCMQ 2254
Qy 96 SGEPTSCPATDAETASGSVESASSTRSGQASSTAVKRPASSEKVKGGDDHDDTSDSD 155
Db 2255 KQENANSTNWTCSRGSVN-----EHALTPSRQTDTEWKIR-----TSLRR 2298
Qy 156 SDGLTLKELQNLRRKREOEPTERPLKGIQSRLEKRRREGPAETV-----GSE 204
Db 2299 PNAMTSSQNFRLKKNRSKNDEVAELGEQKQSLERHKKLLKNILKRSLLERLNQSE 2358
Qy 205 ASDTVEGVLSKQEPENDQGVVQAGKDDR--ESKLEKGAQDIKDHEPGD---LGRPKP 259
Db 2359 IHEDVTKVQVRHVRP-----LSNASPDEQSENERSGEPNLDPRTEVQPRHGAGRPKK 2412
Qy 260 -----ECE-GYD 265
Db 2413 LTRKKEKLYCICRTFYDDTKFYVGCDCLSNWFHGDVCSITEASKLSEFICIDCKRARE 2472
Qy 266 FNALYCICROPHN-NRFNICDCRCEWPHGDCVGISEARGLLERNGEDYICNCTILOV 324
Db 2473 TQQLYCSROPYDESQFYICCDKQDFHGRVCVGLQSEAEFI-----DEVVCPECQ----- 2524
Qy 325 QDETHSETAQOERAKWRPGDAGTDCSTIGTIEQKSSDDQIKGRIEKAA---NPSGKK 380
Db 2525 -----KNDANANMKKLTNDVVELKNLIKQMLHKSAPFPPEVDPK 2568
Qy 381 KLKIFQPVIEAP 392
Db 2569 EAPDYKVIKEP 2580

RESULT 5
Q95VB8 PRELIMINARY; PRT; 2669 AA.
AC Q95VB8;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Nucleosome remodeling factor, large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CGI0894 OR CGI17135 OR CG32346 OR CG32478.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21469388; PubMed=11583616;
RA Xiao H., Sandaltzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
FA Fu D., Wu C.;
ET Dual functions of largest nurf subunit nurf301 in nucleosome sliding
RT and transcription factor interactions.";

RL Mol. Cell 9:531-543(2001).
DR EMBL; AF417921; AAL16644.1; --
DR FlyBase; FBgn000541; E(bx).
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR00637; AT hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300687 MW; 6B4925AFFF49D6F CRC64;

Query Match 6.6%; Score 197; DB 5; Length 2669;
Best Local Similarity 19.0%; Pred. No. 0.0001;
Matches 82; Conservative 75; Mismatches 139; Indels 136; Gaps 15;

Qy 49 EPPPPQOGLSLRRSGRQPKTER-----VEQFLTIARRGRRRMPVSLD----- 95
Db 2217 EHDEPTNLGLDISETDLNKKQNESFVVTGRYIQKSISNALKQ--NLSPELEKLVCMQ 2274
Qy 96 SGEPTSCPATDAETASGSVESASSTRSGQASSTAVKRPASSEKVKGGDDHDDTSDSD 155
Db 2275 KQENANSTNWTCSRGSVN-----EHALTPSRQTDTEWKIR-----TSLRR 2318
Qy 156 SDGLTLKELQNLRRKREOEPTERPLKGIQSRLEKRRREGPAETV-----GSE 204
Db 2319 PNAMTSSQNFRLKKNRSKNDEVAELGEQKQSLERHKKLLKNILKRSLLERLNQSE 2378
Qy 205 ASDTVEGVLSKQEPENDQGVVQAGKDDR--ESKLEKGAQDIKDHEPGD---LGRPKP 259
Db 2379 IHEDVTKVQVRHVRP-----LSNASPDEQSENERSGEPNLDPRTEVQPRHGAGRPKK 2432
Qy 260 -----ECE-GYD 265
Db 2433 LTRKKEKLYCICRTFYDDTKFYVGCDCLSNWFHGDVCSITEASKLSEFICIDCKRARE 2492
Qy 266 FNALYCICROPHN-NRFNICDCRCEWPHGDCVGISEARGLLERNGEDYICNCTILOV 324
Db 2493 TQQLYCSROPYDESQFYICCDKQDFHGRVCVGLQSEAEFI-----DEVVCPECQ----- 2544
Qy 325 QDETHSETAQOERAKWRPGDAGTDCSTIGTIEQKSSDDQIKGRIEKAA---NPSGKK 380
Db 2545 -----KNDANANMKKLTNDVVELKNLIKQMLHKSAPFPPEVDPK 2588
Qy 381 KLKIFQPVIEAP 392
Db 2589 EAPDYKVIKEP 2600

RESULT 6
Q9W0T1 PRELIMINARY; PRT; 2669 AA.
ID Q9W0T1
AC Q9W0T1;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=Berkeley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RC Ananthakrishnan P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.P.,
RX George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A., Fleischmann W.,
RA Fostler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberg J., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananthakrishnan P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Paclet J., Faragas V., Park S., Patel S., Pfeiffer B., Scheeler P.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler P.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL; AE003536; AAF49773.3; -;
DR FlyBase; F59n0036398; C39007.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
SQ SEQUENCE 3146 AA; 336443 MW; 81EEAC5C308FEC31 CRC64;
Query Match 6.5%; Score 194; DB 5; Length 3146;
Best Local Similarity 20.5%; Pred. No. 0.00021;
Matches 121; Conservative 61; Mismatches 235; Indels 174; Gaps 24;
QY 33 TIAKRE--CAGDAEADPLEPPPPPPQOOLGLSLRSRGRQKPTERVEQFLTIARRRERSMP 90
DB 575 TAAKQSVASPSRAVSLERKQHQOQDQVIGGR--KAPTIVIEY----NKHGYNIV 628
QY 91 VLSRSGEPTSPATDAETASGS-----VSSAETRSGPOSASTAVKRPASSSEKVG 144
DB 629 GSSNNLAQNSMSNLAPRSGSGFATTTPTATPLHLTPVNVFVHVEAAPSPSPALVK 688
QY 145 GDHDDTSDSDGLTLKQLNLRKQEPTE--RPLKGIQSLR----- 189
DB 689 GSSPPPAQOQQOQQOQHPGLPQMLNANDELYIEVFPVLTQDLRLOOLHAIMQDHTY 748
QY 190 -KKREBEGPAETVGSSEASDTVEGLVPSKQEP-----NDQVVSQAG----- 230
DB 749 ASQOQQOQQOQQAAG--DTNPGAAQVQOQPOQWSLGGIGVTVSGSQGTPTAVGGYCSYF 805
QY 231 -----KDRRESKLE-----GKAAQIDKBEQDILGRPKPECGYDPNLYCICROP 276
DB 806 GQCIASQADDDAHSASISSRMLASTDI-----DPGETETAPAEARSDSVTRCICELT 862
QY 277 HNNFMICDCRCEENFHGCVGISSEARGLIERNGEDYICPNCITILQVODETHSETADQQ 336
DB 863 HDGTYMCCDKCSANQHVDCWGLDR-----QNIPEYWCELCQPRAV-DKARARALQRQ 915
QY 337 EAKWR-----PGDADGTDCTSTGTI-----FQKSESDQGIKGR----- 370
DB 916 KRKEHMLVATQAANGAAVAAGTTLSGGIAGSLPMSRBLQHLRLASGLNGGPFATGMSK 975
QY 371 -----EKAANPSGKKKLIPOVIRAPGASKICGFCGCHVAQPSVYCSNDILKHAHA 424
DB 976 KSKYKENGSGSTLEKTK--KSAVCGGEGKNSGSG-----TPTG----- 1014
QY 425 TWKPLSSGKEQKPK 484
DB 1015 -----SSGKTSKSKSKSKSKSGD-----GSSGG-----GSSPALTAARKHAANLRQ--- 1056
QY 485 VPARSEALKEACBSSTPSWASDHNY---NAVKEPKTA-----APSPILL 527
DB 1057 -----WIENYATVNHYSPELRLALHAIQKPSLL 1087
RESULT 8
Q727D6 PRELIMINARY; PRT; 2764 AA.
ID Q727D6
AC Q727D6; (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)

01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Bromodomain PHD finger transcription factor.
 OS Homo sapiens (Human)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barak O., Lazzaro M., Lane W., Speicher D., Picketts D.,
 RA Shiekhattar R.;
 RT "Isolation of human NURF";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY282495; AAF22284.1; --
 SQ SEQUENCE 2764 AA; 308079 MW; 9785D7097C2099F9 CRC64;

Query Match 6.4%; Score 191; DB 4; Length 2764;
 Best Local Similarity 21.0%; Pred. No. 0.0029;
 Matches 86; Conservative 58; Mismatches 121; Indels 144; Gaps 18;

QY 35 AKREGAGDAEADP--LEP---PPQQQLGLSLRRSGROPK-RTE-RVEQFLTIARRRR 87
 DB 2245 AQSSSVGAXAQPQTAQSPAXPQPTQ-----PQSPAQPEVQTQDEVQTITVSSHVPS 2299
 QY 88 SMPVSLDSGEP---TSCPATDAETASGVSSESTSGPQSASTAVKRPASSEKVK 143
 DB 2300 AQPTAQ--SSKQVAQAQSPQSNVQSPVVSQSPQIRP---STPSQLSPGQSQVQ 2355
 QY 144 GGDHDDTSDSP-----SDGLTLKELQ-----NRLRRKREQEETERP- 180
 DB 2356 -----TTTSQPIPIQHTSIQIPSQSQPQVPMKHNVAIBHLKQKMSMTFAEREE 2408
 QY 181 -----LKG1-----QSLRKKRRREGPAETVGSSEADTVGVLPSKQSPENDQ 223
 DB 2409 NORMIVCQVMKYLIDKDKXEQAKKRRSEVQKESQNAKLSALLFKHKEQLRA 2468
 QY 224 GYVSQAQKDDRESKLE--GKAADIKDEPGDL-----GRPPECEG 263
 DB 2469 EILKKRALDKLOIEVQSELRKDLKIKKEDLMQLAQTAVAPCPVTPAPPAPP 2528
 QY 264 YDP-----NLYYC 271
 DB 2529 PSEPPPPAVQHTGLSTPTLPAASQVRKREERKSSSKKKKMTSTTSKETTKDTLYC 2588
 QY 272 ICRQPHN-NRFMICDRCRWFHGDVGVISAGRLLENGSDYICPNC 319
 DB 2589 ICKTPVDESKEFYIGDRCQWYHGRVGVILQSEAEILI-----DEVVCPQC 2633

RESULT 9
 Q818F0 PRELIMINARY; PRT; 542 AA.
 AC Q818F0;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Chromosome scaffold protein p85.
 OS Euplotes crassus.
 CC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
 CC Euplotida; Euplotidae; Moneuplotes.
 ON NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sharp S.I., Pickett J.K., Jahn C.L.;
 RT "The identification of a novel chromosome scaffold protein that
 RT associates with rec elements undergoing en masse elimination in
 RT Euplotes crassus";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY155457; AAN75020.1; -- F0A3B5E3AA771A1 CRC64;
 SQ SEQUENCE 542 AA; 60264 MW; 60264 MW; 9785D7097C2099F9 CRC64;

Query Match 6.3%; Score 187.5; DB 5; Length 542;
 Best Local Similarity 19.4%; Pred. No. 7.1e-05;
 Matches 110; Conservative 94; Mismatches 189; Indels 175; Gaps 23;

QY 6 DPNEZAPKAIKPTSKPRKWTGFRRTTIARREGAGDAEADPLEPPPPQQQLGLSLRRSG 65
 DB 42 NPSDEE-----DINKESAFSTHTNKSSEKSDVEENNDDHA-----PKQEVKTRKTSS 88
 QY 66 RQ-----PKRTERVEQFLTIARRRGRSMPVSLDSGSEPTSCPATDAETASGVSSEASE- 120
 DB 89 RKTAPPKTKTKENK--VSGKTRKVKKEPDDE-----EVSEDKNVAKSSSI 135
 QY 121 -TRSGPOSASTAVKRPASSEKVKGGDDHDDTSDSDSLTLKELQNLRLRRKREQEETER 179
 DB 136 LRSSRLAAANANKLSAK---KG--ENQNSDSEYEEDTPK-----SKRQAKKASK 183
 QY 180 PLKGIQSLRLKRRREGPAETVGSSEADTVGVL--PSKQBPENDQGVVSQAQKDDRESKL 238
 DB 184 DESSSEAESESESEKSPKSKSKATKPAKVVKPTK-----AAKTAKTKKI 231
 QY 239 EGKAAQDIKDEPGDLGRPCEGYPDPNLYCICROPHNRFMICDRCRWFHGDVGV 298
 DB 232 E-----EDDEPSDESQSGEGEQEEA-----KSQDSQK-----D 262
 QY 299 ISEARGLLRNGRDYICPNTILQVODETHSETADQOEAKWRPGDADTCTSIGTIEQ 358
 DB 263 SGEGEG--HENDP-----EMEEAEGEEQDEDEMSSEAEGBE-----EE 302
 QY 359 KSSQDQIKGRIEKAANPSGKKLKIPOPVIEAPGASKICIGFCCHVAQPSVVCSDNCI 418
 DB 303 TVKDEGKKGKVEBEKKVAGKK-----DSPKKKK-----DSPKKKKDSP 342
 QY 419 LKHAATAKTLSSCKEQKPK-----P 439
 DB 343 GKEDVKKKVTSEKKNKSPDKKEAAKKEKVAKDQATKKKDKTEKKKEITKQDAT 402
 QY 440 KEKMKPEKPSLPFCGAQAGIKISSVHKRPAPKEKTTVKKAVVVPARSALGKEAAACE 499
 DB 403 KDKDIQKEKTSFKDSKAGTK--EHKRASQSEKKKDVKS---PKDKATSCKGKAE 456
 QY 500 S-----STPSWASDNYNAVPEKTAAPS 523
 DB 457 SKKKDSFKDKKKDKKISPKNGKASS 484

RESULT 10
 Q8C969 PRELIMINARY; PRT; 443 AA.
 ID Q8C969;
 AC Q8C969;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN A630082K2ORIK.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=1246851;
 RA The FANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL: AK042834; BAC31377.1; --
 DR MGD; MGI:2443388; A630082K2ORIK.
 DR GO: GO:0003677; F:DNA binding; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR003347; TF Jm1C.
 DR InterPro: IPR001965; ZnF_PHD.
 DR Pfam: PF02373; Jm1C; 1.
 DR Pfam: PF00628; PHD; 1.

DR SMART; SM00558; Jm1C; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS00016; ZF_PHD_2; 1.
 KW Hypothetical protein.
 FT NON_TER 443 443
 SQ SEQUENCE 443 AA; 49944 MW; C543F487717C4682 CRC64;
 Query Match 6.3%; Score 187; DB 11; Length 443;
 Best Local Similarity 22.0%; Pred. No. 6e-05;
 Matches 78; Conservative 44; Mismatches 118; Indels 114; Gaps 18;
 QY 240 GKAAQDIKDEBGLGRPKPCBCEGTDNALYICICRQPHN-NRPMICCDRCCEWFGHDCVVG 298
 Db 15 GAAAAGSAGSAPGRASAPPP-----PPVYVCVQPYDVNRFMIECDVCKWFGHSCVVG 68
 QY 299 ISARGELLERNGEDYICPNCTILOVDETHSETAQOEAQKWRPGD-----ADGTDCTSIG 354
 Db 69 VEEHHAVIDL-----YHCPDCAAL-----HGSLMKKRRNHRHDYTEVDDGSKPVOAG 118
 QY 355 T-----IRKSSDQD--IKGRIEKAANPSGKKLKIIFQ 386
 Db 119 TRAFVKELSRVFPSSADEIIVKMGHSQLTQRYLEKHGFDVPIVMPKLDLGLG-----LRLPS 174
 QY 387 PVIEAPGASKICPGCGCHVAQPDVYCSNDICLKHAATAWKFLSSGKEQKPKKEMKMK 446
 Db 175 PAFSVMQVRYVG-----GDKVIIDVID-VARQADSKTLEN-----YVKYFMN 216
 QY 447 PEKPSLPKCGAAGIKISSVHKRPAPKPKETTIVKAVVVPARSEALGKEACSSSTPSWA 506
 Db 217 PDRPKV-----LNVISL-----EFSDTKMSLVEVFDIARKL-----SWV 251
 QY 507 SDNYNAVPEKTAAPSPSLLYKCM-----YHLGVGLLDPSPRSFWIAIPW 551
 Db 252 --ENY---WPDDSVFPKPPVQKCYCLMGVQDSYTDHFIDFG-----GTSVWYHVLW 296
 RESULT 11
 Q8C9E0 PRELIMINARY; PRT; 473 AA.
 ID Q8C9E0
 AC Q8C9E0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (fragment).
 GN A630082K20RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466653;
 RA THE FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RL EMBL; AK042327; BAC1286.1;
 DR MGD; MGI:2443388; A630082K20RIK.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR003347; TF Jm1C.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF02373; Jm1C.1.
 DR Pfam; PF00628; PHD; 1.
 DR SMART; SM00558; Jm1C; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.

DR PROSITE; PS00016; ZF_PHD_2; 1.
 KW Hypothetical protein.
 FT NON_TER 473 473
 SQ SEQUENCE 473 AA; 53409 MW; 94C94378609C7039 CRC64;
 Query Match 6.3%; Score 187; DB 11; Length 473;
 Best Local Similarity 22.0%; Pred. No. 6.5e-05;
 Matches 78; Conservative 44; Mismatches 118; Indels 114; Gaps 18;
 QY 240 GKAAQDIKDEBGLGRPKPCBCEGTDNALYICICRQPHN-NRPMICCDRCCEWFGHDCVVG 298
 Db 15 GAAAAGSAGSAPGRASAPPP-----PPVYVCVQPYDVNRFMIECDVCKWFGHSCVVG 68
 QY 299 ISARGELLERNGEDYICPNCTILOVDETHSETAQOEAQKWRPGD-----ADGTDCTSIG 354
 Db 69 VEEHHAVIDL-----YHCPDCAAL-----HGSLMKKRRNHRHDYTEVDDGSKPVOAG 118
 QY 355 T-----IRKSSDQD--IKGRIEKAANPSGKKLKIIFQ 386
 Db 119 TRAFVKELSRVFPSSADEIIVKMGHSQLTQRYLEKHGFDVPIVMPKLDLGLG-----LRLPS 174
 QY 387 PVIEAPGASKICPGCGCHVAQPDVYCSNDICLKHAATAWKFLSSGKEQKPKKEMKMK 446
 Db 175 PAFSVMQVRYVG-----GDKVIIDVID-VARQADSKTLEN-----YVKYFMN 216
 QY 447 PEKPSLPKCGAAGIKISSVHKRPAPKPKETTIVKAVVVPARSEALGKEACSSSTPSWA 506
 Db 217 PDRPKV-----LNVISL-----EFSDTKMSLVEVFDIARKL-----SWV 251
 QY 507 SDNYNAVPEKTAAPSPSLLYKCM-----YHLGVGLLDPSPRSFWIAIPW 551
 Db 252 --ENY---WPDDSVFPKPPVQKCYCLMGVQDSYTDHFIDFG-----GTSVWYHVLW 296
 RESULT 12
 Q95YM2 PRELIMINARY; PRT; 17352 AA.
 ID Q95YM2
 AC Q95YM2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE I-connectin.
 GN I-CON.
 OS Procambarus clarkii (Red swamp crayfish).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
 OC Astacoidea; Cambaridae; Procambarus.
 OX NCBI_TaxID=6728;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21423462; PubMed=11532946;
 RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
 RA Sun P., Maruyama K., Kimura S.;
 RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
 RT sarcomeres of crayfish claw muscle.";
 RL EMBO J. 20:4826-4835(2001).
 CC I-SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AB055861; BAB64297.1;
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR000577; FGGY kin.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig c2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; Ig; 49.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IGG2; 13.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00933; EGGY_KINASES_1; 3.
 DR PROSITE; PS00835; IGLIKE; 49.
 DR PROSITE; PS00002; SH3; 1.

KW Immunoglobulin domain; SH3 domain. 4BA157BEC042E42D CRC64;
SQ SEQUENCE 17352 AA; 1962348 MW; 2.3%; Score 187; DB 5; Length 17352;
Query Match 6.1%; Score 182; DB 13; Length 563;
Best Local Similarity 21.9%; Pred. No. 0.00018;
Matches 68; Conservative 31; Mismatches 66; Indels 146; Gaps 11;
Matches 136; Conservative 81; Mismatches 206; Indels 218; Gaps 32;
QY 2 DKGDPSENE--APKAIPTSKFRKTGWRRTTIKREGAGD-----AEADPLEP- 50
DB DEEKLPSDEQKLKKAQVKKDBEEIPSWGRKLPPEDEKEEITLKPKKVKPKPK 6267
QY 51 PPPQOQLGLSLRRSGRQKRTVERVEQFLTIARRGRSRMPVSLSDSEFTSCPATDA 110
DB PSPKLPKPKYEPPIPEPKS-----PL-----EPYSKPEKE--V 6300
QY 111 SGVSFESAFTRSGPQASSTAVKPPASSEKVKGGDDHDDTSDSDGLTTLKELQRL 170
DB 6301 SDKIPELAKEVSRPKEE-----PEKPEPEKL-----DESDKKPDESETIT 6348
QY 171 KREQPTERPLKGIQSLRLKRR-----EEGPAETVSGEASDTVEGVLPFS----- 215
DB 6349 QKLKAPTSEKEIPKVTLRKTSQKVPVPEVLTETVLEHVEITPEWPEVEKRWSP 6408
QY 216 -----KQEPNDGVVSOAGKD-----DRESKL- 238
DB 6409 EYETVVPSEIPEKPEVELEKYEKYPPTKPKDDEEDKGYERKPKPKPEPDRK 6468
QY 239 ESKAAQDIKDBEPGD-----LGRPK-PECEGYDPNALYCICROPHNNRPMICD 290
DB 6469 KKK-----LRPEEGEKKLKKPKRPKSPKPEAKXPQLKPKPKPE----- 6511
QY 291 WPHGDCVGISEARGLLRNGDYICPNCTI-----LOVODETHSETAQOEA----- 338
DB 6512 -----EKKQKVTPKPGKK-----PSKKIPDRFVEPEPFEPTPEBILDK 6559
QY 339 ---KWRGDADGTDCTISGTHQKSSDQIGKRIEKAANSGKKLKIFQPVIAFG-A 394
DB 6560 LKPKPEPEKPSIPEAPKPELEKESSEEPKPKKER---PKPEK---BEAEVPS 6612
QY 395 SKCIGPGCHVAQPSVYCSNDCILKHAAMTKFLSSGKQKPKPKPKPKM-KPKPS 453
DB 6613 GKRLPP---KEEKEEI-----VLKPKPKPEPEPKPKPKPKPKPKPKPK 6656
QY 454 K-----CGAAGIKISSVHKRPAPKPKETTVKKAIVVPAARSALGKE 495
DB 6657 EPEKTPLEPYTKPKREKVPDGVTEPVKPED-SEKPKPE-BEIKPKKERIKPK 6709
QY 496 AACESGTSW-----ASDHNVAVKP-----EKTAAPS 525
DB 6710 ---EVETPSWGRKLPPEKDEKKEEITLKPKPKPKPKPKPS 6747
PRLIMINARY; PRT; 563 AA.
RESULT 13
Q7SZX6
ID Q7SZX6 PRELIMINARY; PRT; 563 AA.
AC Q7SZX6; (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE SI:zk13a21.10 (Novel protein similar to human and mouse Cpg binding protein (CGBP)).
GN SI:zk13a21.10.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pandian R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844847; CAC30421.1; -- A084ED9C4914BB83 CRC64;
SQ SEQUENCE 563 AA; 66601 MW; A084ED9C4914BB83 CRC64;

Query Match 6.1%; Score 182; DB 13; Length 563;
Best Local Similarity 21.9%; Pred. No. 0.00018;
Matches 68; Conservative 31; Mismatches 66; Indels 146; Gaps 11;
QY 248 DEBFG-DLGRPKPECEGYDPNA-LYCICROPHNNRPMICDRCERHFWHFCVGLSE----- 301
DB 9 DQTEGLDMSKEGE-----NAPLYCTCRKSDINCWFGDCNBNWPHGHCINVTQAK 62
QY 302 -----ARGRLLRN----- 310
DB 63 AIREWCOQCRARDPSLSIRYKKNRDKDVEPERVEKRSSTPEYKIDKRGSKVKRSARM 122
QY 311 -GEDIYICPNT-----LOVODE----- 327
DB 123 CGE---CEPCTRTEDCGHCFCKDMKFGGPNKIQRKLRQCVVRRKMLRVDEEFS 179
QY 328 -----THSETAQOEB-----AKWRPGDADGTDCTSIGTIEOKSSEDQ 364
DB 180 RERKDNIMHDDRYSDYDNDMDLYHYKORNASMGSEDDQGLI9PVP--RKAIVK 237
QY 365 GIKRIEKAANPSGKKLKIFQPVIE-----AFGASKCIGPGCHVAQPSVY 412
DB 238 HVKEEDKFKKESRRHKQKQKRDRLRHSRDRDGRHGSDTQCLGPNCTIEPARPNSKY 297
QY 413 CSNDCLIKHAA 423
DB 298 CSEDCGMKLA 308
PRLIMINARY; PRT; 563 AA.
RESULT 14
Q7SZE4
ID Q7SZE4 PRELIMINARY; PRT; 563 AA.
AC Q7SZE4; (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AB; TISSUE=Body;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W., Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.S., Scherch A., Schein J.E., Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052918; AAH52918.1; --

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:51:47 ; Search time 21 Seconds
(without alignments)
2574.268 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKEDPSREAPKAIKPTS.....RSFWIAIPWACPLGVAAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:**
1: Pir1:**
2: Pir2:**
3: Pir3:**
4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	6.5	488	2 I46014	Cylicin II - bovin
2	189.5	6.3	706	2 A45990	junctional sarcopl
3	178	6.0	606	2 A43427	neurofilament trip
4	175	5.9	424	2 T41449	probable phd frige
5	175	5.9	5327	2 T13564	microtubule-associ
6	174.5	5.8	669	2 J05662	hepatoma-derived g
7	171.5	5.7	699	2 I38073	nucleolar phosphop
8	171.5	5.7	1110	2 I31116	NF-180 - sea lampr
9	170	5.7	510	2 T21430	hypothetical prote
10	169	5.7	729	2 S68191	triadin - human
11	169	5.7	990	2 I51618	nucleolar phosphop
12	167.5	5.6	1684	2 J00057	gravin - human
13	166	5.6	452	2 T21435	hypothetical prote
14	165.5	5.5	734	2 B42680	nucleolus-cytoplas
15	165.5	5.5	1057	2 H83273	ribonuclease E pA2
16	164	5.5	1359	2 T34036	hypothetical prote
17	159.5	5.3	1087	1 QPM5H	neurofilament trip
18	157.5	5.3	405	2 T21433	hypothetical prote
19	157	5.3	1173	2 T31421	C-terminal domain-
20	156	5.2	1298	2 I54367	X-linked nuclear p
21	156	5.2	1840	2 T30250	GTL protein - mous
22	155.5	5.2	1560	2 T42727	proliferation pote
23	154.5	5.2	1588	2 T38660	probable transcrip
24	154.5	5.2	1830	2 A37981	microtubule-associ
25	153.5	5.1	1200	2 A46194	neurofilament prot
26	153.5	5.1	1233	2 T41457	serine/threonine p
27	153.5	5.1	3488	2 T34418	hypothetical prote
28	153.5	5.1	6642	2 T29757	protein UNC-89 - C
29	153	5.1	952	2 T18837	hypothetical prote

30	152.5	5.1	1641	2 I38614	helicase II - huma
31	152	5.1	763	2 T08929	hypothetical prote
32	152	5.1	1020	1 QFHUH	neurofilament trip
33	151.5	5.1	849	2 S00030	neurofilament trip
34	151.5	5.1	856	2 T16543	hypothetical prote
35	150.5	5.0	272	2 T02745	nucleic acid bindi
36	150	5.0	713	2 A28706	calpastatin, card
37	150	5.0	771	1 A33430	h-caldesmon - chic
38	150	5.0	1171	2 T13065	PIP2 protein - fr
39	149.5	5.0	644	2 S55395	neurofilament prot
40	149.5	5.0	902	2 T47966	hypothetical prote
41	149.5	5.0	1622	2 J80378	DNA (cytosine-5)-
42	149	5.0	273	2 T51145	nucleic acid bindi
43	149	5.0	810	2 T44430	protein PVI00 (imp
44	148.5	5.0	679	2 S28366	recombination repa
45	148.5	5.0	1560	2 I54361	SMCX protein - hum

ALIGNMENTS

RESULT 1

I46014
Cylicin II - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-Nov-1999
C:Accession: I46014; S52772
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.
Exp. Cell Res. 218, 174-182, 1995
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the
A:Reference number: I37271; MUID:95255491; PMID:7737358
A:Accession: I46014
A:Status: preliminary; translated from GB/EMBL/DDBV
A:Molecule type: mRNA
A:Residues: 1-488 <RES>
A:Cross-references: EMBL:Z46789; NID:g757753; PIDN:CRA86753.1; PID:g757754

Query Match 6.5%; Score 193; DB 2; Length 488;

Best Local Similarity 20.3%; Pred. No. 0.00064;

Matches 117; Conservative 73; Mismatches 204; Indels 182; Gaps 19;

Qy	18	PTSKEPKTWGPRRTIAKREGAGDAEADPLEPPPPQQLGLSLRRSGRQPKRTERVEQF	77
Db	19	PVSELSKSN-----QQHFALVPKPPPKR-----	46
Qy	78	LTIARRRRGRSMFVSLSDSGEPTSCPATDAE-----TASEGSVESAS	119
Db	47	-----RSKPSLLQEN-----TSPKYDAEKLGRDKQPLMWHRSLSRISRPSVYLAA	94
Qy	120	ETR-----SGPQSGASTAVKERPASSKVKGGDDHDDTSDSDGLTKELQNLRRKEEQE	175
Db	95	RSRHPQKTEPPSOEDAKQAAPSPKVKSKEDKSDSEARSIVSKEKPKKSKAKEEK	154
Qy	176	PTER-----PLKGIQS-----RLRKKRRRGGPAFTVVGSEASDT	208
Db	155	PDEKDLKKERKSKKGSATSEDEKAGAKGAKDKGSKKGTPTSDS--GSEKGA	213
Qy	209	VEGVLPSPKQPE-----NDQGVVSQAGKDDRESK-----LEGKAAQDIQDEE	250
Db	214	KDSKSKSKGSATSEGEKGDADKDDKKGKSGSKGKESATESGEKGAQKDDK	273
Qy	251	PGDLGRPK-----PECRGYPNALYCICRQPHNNRPMICCDRCCEWPHGDCVIGSEARGR	305
Db	274	KGKSGKSGKGSATESGEKGDAAK-----DDKKGK	304
Qy	306	LLERNGEDYICPNCTILQVQDETHSETADQGEAKWRPGDADGTCTCTGTYEQKSSDEQ	365
Db	305	KGSKKGE-----SATESGEKGDADKDDKKGKSGKSGK-----KESATESG	347
Qy	366	IKGEIEKAAPSGKKLKIQQPIEAPGASKIGEGCHVAQPSVYCSNDILKHAAT	425
Db	348	ERGDAAK--DDKKGKK-----GSKK--GKESDSKAEKGDGAKKD--DKDKKGS	391

Qy	426	MRFSLSGKQKPKPKPKMKMKPKESLPKCGAAGIKLSSVHKPAPKPKETTVKKAIVV	485
Db	392	KKGKGAATESGKKDKSKDKRAGKDPPTKAGEKD---ESKKDKAKKDKSKKDKKK	448
Qy	486	PARSEALGKRAACESTPTSPASDHNYNAVPEKTA	521
Db	449	PGEASEPKDSAKKDKAKKDKAKKDKAKKDKAKKDKA	484

RESULT 2

A45990
 Junctional sarcoplasmic reticulum glycoprotein triadin - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 11-Jan-2000
 C:Accession: A45990; S45704
 R:Knudson, C.M.; Stang, K.K.; Moonaw, C.R.; Slaughter, C.A.; Campbell, K.P.
 J. Biol. Chem. 268, 12646-12654, 1993
 A:Title: Primary structure and topological analysis of a skeletal muscle-specific junctional sarcoplasmic reticulum glycoprotein triadin from rabbit
 A:reference number: A45990; MUID:93286104; PMID:7685347
 A:Accession: A45990
 A:Status: preliminary
 A:Molecule type: mRNA; protein
 A:Residues: 1-706 <KNU>
 A:Cross-references: GS:I10065; NID:G347850; PID:AAA31488.1; PID:G347851
 A:Experimental source: triads, muscle
 A:Note: sequence extracted from NCBI backbone (NCBIN:133603, NCBIPI:133604)
 R:Peng, M.; Pan, H.; Kirely, T.L.; Caswell, A.H.; Schwartz, A.
 FEBS Lett. 348, 17-20, 1994
 A:Title: Structural diversity of triadin in skeletal muscle and evidence of its existend
 A:reference number: S45704; MUID:94289946; PMID:8026576
 A:Accession: S45704
 A:Molecule type: mRNA
 A:Residues: 401-620 <PEN>
 C:Superfamily: histone H1
 C:keywords: glycoprotein

Query Match	6.38;	Score 189.5;	DB 2;	Length 706;
Best Local Similarity	22.98;	Pred. No. 0.0015;		
Matches	99;	Conservative 62;	Mismatches 177;	Indels 95; Gaps 16;
QY	145	GDHDDTSDSDGLTKE--LQNLRRKREQEPTERPLKGIQSL--	--RKKRREGPAAE	199
DB	118	GDEDEDEGTAKGEIEEPPLKEKDIHKKEIEKQEKPERKIPTKVHHEKEKEKEKVKKE	177	
QY	200	TVGSASATVTEGVLPSEQEPENDGVVQAGKDDRESKLEKGAQAQDIKDSEPGDLGR---	256	
DB	178	KEKPEKGAATHKEKLEKKEKET-KTVTKEEKKARTKEKTEETKEKVGAKVQEKVQTVA	236	
QY	257	-----PKP-ECGEYDNLALYICICQPHNNRPMICCDRCBEWFHSGVCGVISEARGELL	307	
DB	237	KAKEVQKTPKPEKESKETAA--VSKQEQDQYAFCRYMIDIFVGD-----L	282	
QY	308	ENGEDYTCNMCTILOVODETHS-ETADQQAkWREGDADGDTCTSIGTIFQKSSHQGI	366	
DB	283	KFGOSPATPPSPPTQASRTPPALPTPEKEGEKKAAE-----KKVTTETKKAAREDA	336	
QY	367	KGRTEKAANPSGKKKLAKI FQPVIEAPGASCKIGRGCHVAQPDVSYCSNDCILKHAATM	426	
DB	337	KKSEKETDIDMKK-----EPKSPDTPKGTGVVTT-----QAATK	373	
QY	427	---KPLSGGKQKP--KPKENMKRPEKPSLPKGA-----QAGIXI-----	463	
DB	374	KDEKEDSKAKAKPAEBQPKGKQKEKKEHEEPKSTKKEHAAPSEKQAKAKIERKEBVS	433	
QY	464	SSVHKRPAPPEKETTAKVAVVAPVARSALCKEAAC-----ESSTPSWASHDNYNA	513	
DB	434	AASTKQAVPAKBEKTKTVEQTRKEKPKGISVVLNDKELTKKEKVKVPSLKEKGSSET	493	
QY	514	VKPEKTAAPSPSL	526	
DB	494	KQDEKTSKPEPOI	506	

RESULT 3

A43427
neurofilament triplet H1 protein - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1999 #text_change 10-Dec-1999
C:Accession: A43427
R:Soppet, D.R.; Beasley, L.L.; Willard, M.B.
J. Biol. Chem. 267, 17354-17361, 1992
A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polypeptide
A:Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBIP:112011)
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.0%; Score 178; DB 2; Length 606;
Best Local Similarity 19.9%; Pred. No. 0.0057;
Matches 100; Conservative 67; Mismatches 211; Indels 124; Gaps 17;

QY 35 AKREGAGDAADPLEPPPPQQLGLSRSGRQPKRTVERVEQFLTIARRRRRSMVSLR 94
DB 28 AKHEEGGEZ---EAKSTEGCAASPREEAKSPA-EK-----SPVK-E 66
QY 95 DSGETPTCPATDAETASGEVSASSETRSGQSASTAVKRPASSEKVKGGDDHDDTSDS 154
DB 67 EAKSP-----AEAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 118
QY 155 DSDGLTLKELQNLRLAKRKEOEPTERPLKGIQSLRKRREEGPAETVGSE-ASDTEGVTL 213
DB 119 EA-----KSP-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 172
QY 214 PSK-----OREPNDGGVVSQAGKDDRESKLEGAQDIDKEPGLGRPKPECGY 264
DB 173 PEKAKSPVKEEAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 227
QY 265 DFNALYCIQEPHNRFMICCDRCCEWTFHGDGCVGTSARGRLRLRMGEDYICPNCITLQV 324
DB 228 -----AEAKSPVKE-----EA 238
QY 325 QRETHSETADQOEAKWRGADAGTCTSTGTTQKSSSDGGIKGRIRKRNPSGKKLKI 384
DB 239 KSPEKAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 288
QY 385 FQPV-IEAPGASKCTGPGCHVAQPDVSVCSNDCLIKUAAATMTLSSGKEQKXP-KEK 442
DB 289 KSPVKEEAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 345
QY 443 MCKEKEKPSLPKCGAAGIKTSVHKRPAPEKKTITVKAVVPARSEALGHEACEST 502
DB 346 SPVKEEAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 402
QY 503 PSWASHDNNYNAVPKETAAASP 524
DB 403 PE-----KAKSPVKEEAKSP 417

RESULT 4
T41449
probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C:Accession: T41449
R:Rieger, M.; Lynne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z11994
A:Accession: T41449
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-424 <RES>

A;Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05c
A;Experimental source: strain 972h-; cosmid c594
C;Genetics:
A;Gene: SPDB:SPCC594.05c
A;Map position: 3

Query Match 5.9%; Score 175; DB 2; Length 424;
Best Local Similarity 21.3%; Pred. No. 0.0058;
Matches 79; Conservative 37; Mismatches 77; Indels 178; Gaps 15;
QY 91 VSLDSGPTSCPATD--AFTASGVSASSTRSQSASTAVKRPASSEKVGKGDH 148
DB 20 VTFEDSNRGT---ITDFHETANNEBEKDAVTLN-----KSVMEVEEVNG---H 64
QY 149 DWTSDSDSLGLTLKELQNLRLRRRREKQSPTEPLKGIQSLRKKRRRREKGPATVGSASDT 208
DB 65 VDSSTETD-----TEMVIOQPTIPKPPVSAHRGRKRKRNA----- 104
QY 209 VEGVLPKSKOBENDQGVVSQAGKDDRESKLEGAQDIKDBEPDGLGRPKCEGYDPNA 268
DB 105 -----NSQLNLSTA-----DHQRP----- 118
QY 269 LYICICPHNRRPMICCDRCBEPFHGDCVCSISARGRLLRNGEDYICPNTILQVQDET 328
DB 119 LYICICQKPDGSMWLGDCGDEWPHGTCVNPISYNDLTQV---YFCPKCT----- 166
QY 329 HSETADQOEAKWRPGADGTCTSIGTIEQKSSDQGI-----KGRIEKAANPSGKKLK 383
DB 167 -----BEKGITTKRKKRLRECSNFT----- 188
QY 384 IFOPVIEAPGASKICIGPCHVQAPDSVYCSNDCILKHAATMKFLSSGKEQKPKCKM 443
DB 189 -----RPNNSNYCSD-----KHG---VDFF-----REKV 208
QY 444 KMKPEKPSLPK 454
DB 209 KISTVEPSAIK 219

RESULT 5
Tl3564
Microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
X;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: Tl3564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: 217689
A;Accession: Tl3564
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 5.9%; Score 175; DB 2; Length 5327;
Best Local Similarity 20.7%; Pred. No. 0.082;
Matches 131; Conservative 81; Mismatches 254; Indels 166; Gaps 25;
QY 2 DDKGDPNNEAPKAIKPTSKERKTKWGFRTTIARREGAGDAEADPLEPPPPQOOLGL 59
DB 2937 DEKASRRSVAESVKP-----SSKDATSAPPKSHRSRPSVIG- 2976
QY 60 SLRSRGPKETER-----VEOPLTIARRGR-RSMFVSLDSGEP---TSCPATD 106
DB 2977 SKDEG---DKTTSRVVADSIKESLLVQASRPSEAESLKDAAPQETSRSPEV 3034

QY 107 AFTASGVSASSTRSQ-----QSASTAVKRPASSEKVGKGDHDDTSD--SDSD 157
DB 3035 TESVKGKSPVASKASRPASVAENAKDSADESKQRPESLPQSRAGSIKDEKSPASKD 3094
QY 158 GLTLKELQNLRLRREKQEP-----TERPLKGIQS---RLKRRREKGP-----A 198
DB 3095 EAEKSKESRRSVAEQPFLVSKVSRPASVAESVKDEAKSKESPLMSKEASRPASVA 3154
QY 199 ETVGSRASITV-----GVLPFSKO---EPEN-DQGVVSQAGKDDRESKLEGA--- 242
DB 3155 GSVKDEAKSKESRRSVAESKPLSPKASRPASVAESVKDBADKSKESRRSVAES 3214
QY 243 -----AQDIKDB-----KPDILGRPKPECEGYDPNALYICICQPHNN 279
DB 3215 PLASKEASRPASVAESIKDEAKSKESRRSVAESKPLSPKASRPASVAESVQDE 3264
QY 280 RFMVICDRCBEPFHGDCVCSISARGRLLRNGEDYIC-----PNTILQVQDE 327
DB 3265 -----AKSVDEAKSKESRRSVAESKPLASKEASRPASVAESVQDE 3308
QY 328 T-----HSETADQOEAKWRPGADGTCTSIGTIEQKSSDQGIKRIEKA--ANPS 377
DB 3309 AEKSKESRRSVAESKPLASKEASRPASVAESIKDEAKSKESRRSVAESKPLASKE 3368
QY 378 GKXKLKIFOPVIEAPGASKICIGPCHVQAPDSVYCSNDCILKHAATMKFLSSGKEQK 437
DB 3369 ASRPTSVASVKDEAKSK-----EESRRSVAESKPLASKEASRPASVAESVQDEAB 3421
QY 438 KPKEKRMK--PEKPSLPKGAQAGIKISSVHKRPAPKPTTVKKAIVVPAEALGKE 495
DB 3422 KSKESRRSVAESKPLASKEASRPASVAESVKDBADKSKESRRSVAES--EKSPLASKE 3479
QY 496 AACESTPWSADH---NYNAVPEKTAAPSP 524
DB 3480 ASRPASVAESVKDEAKSKESRRSVAESKSP 3511

RESULT 6
JC5662
hepatoma-derived growth factor-related protein 2 - mouse
N;Alternate names: HRP-2
C;Species: Mus musculus (house mouse)
C;Date: 17-Nov-1997 #sequence_revision 17-Nov-1997 #text_change 05-Nov-1999
C;Accession: JC5662
R;Izumoto, Y.; Kuroda, T.; Harada, H.; Kishimoto, T.; Nakamura, H.
Biochem. Biophys. Res. Commun. 238, 26-32, 1997
A;Title: Hepatoma-derived growth factor belongs to a gene family in mice showing signifi
A;Reference number: JC5660; M01D:97445118; PMID:9299445
A;Accession: JC5662
A;Molecule type: mRNA
A;Residues: 1-669 <IZU>
A;Cross-references: DDBJ:D63850; MID:g2558500; PIDN:EAA22896.1; PID:d1023766; PID:g25585
C;Comment: This protein translocates to the nucleus and directly functions in mitogenic
F;1-93/Region: hath #status predicted
F;321-363/Region: nuclear location signal

Query Match 5.8%; Score 174.5; DB 2; Length 659;
Best Local Similarity 20.9%; Pred. No. 0.01;
Matches 128; Conservative 77; Mismatches 182; Indels 225; Gaps 29;
QY 2 DDKGDPNNEAPKAIKPTSKERKTKWGFRTTIARREGAGDAEADPLEPPPPQOOLGLSL 61
DB 187 EDESPPSEB-----KTSDDQF-----TPEKTA-----RPRRGLG--- 220
QY 62 RRGQRPKTRVERVEQFTIARRREGSRMPVSLDSGEPSTSCPATDAETASGVSASSET 121
DB 221 ---GRKKKVPASDSKADSDGAKPEPV---TAQPS---PSSSSSSSSSSSDSDSVSV 272
QY 122 RSPQSNASTAVK-----ERPASSEKVGKGDHDDTSDSDSLTKELQN--- 166
DB 273 KCPGRGKPAEKPFPKPRGRPRKPPST-----SSSDSDSDSDSEVDRISTWKR 322
QY 167 -----RLRRKREQEPTEPLKGIQSLRKKRRREKGPATVGSASDTVEGVLPFSK 217

Db 323 RDEERRELEARRREQEELRLRE-QREERKERE-RAERGS-----SCE 369
Qy 218 EENDQGVVQAGKDDRESKLGKAAQDIKDEP-GDLGR-----PKPECGYDPNALY 270
Db 370 ELEDEEPV-----KKRSKARGTTPSSSDPEPGLGKEKKLAKSOLPGSES-----419
Qy 271 CFCROPHNNRFMICCDRCBFWHGDVCGISARGLLRNGEDYICPNCTILOVDETHS 330
Db 420 -ARKP-----GQKRRGR-----431
Qy 331 ETADQOEAKWRPGDADGTDCTSIGTIEOKSSDQIGKRIEKAANPSGKKLK-----I 384
Db 432 ---PDEKPARPVKERT-----KKRSGLSLERKGEKKPSVEERLQKLHSEIK 479
Qy 385 FQPVTEAPGASKICIGP---GCHFA-----QPDGV-----YCSNDCILKHA--423
Db 480 FALKVDNPDVRCLSALELGTQVTSQILQKNTDVATLKIRRYKAKDVMAKAARVY 539
Qy 424 -----ATWKFLSSGKEQKPKPKKWKMPKPSLPKCGAQAQIKISSVH-----467
Db 540 TRKSRVLGPKVEALQKVNKAKEKADNKKLEBP-----GEOAPRELAEDPEST 591
Qy 468 KPAPEKKTIVKAAVVPARGAALGKZAAECSSSTPSWASDH-----NNNAVKP-----516
Db 592 DRAPVNGEATQKGNMEDRAQEDQDS--EDGPRGSSSEELHSDPSNDSPPAKPGNER 649
Qy 517 ---EKTAAPSPS 525
Db 650 QDHERTRLASES 661

RESULT 7

I38073
nucleolar phosphoprotein pl30 - human
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 24-Sep-1999
C:Accession: I38073; S52292
R:Pat C.Y.; Chen, H.K.; Shen, H.L.; Yeh, N.H.
J. Cell Sci. 108, 1911-1920, 1995
A:Title: Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein P
A:Reference number: I38073; MUID:95386590; PMID:7657714
A:Accession: I38073
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-699 <RES>
A:Cross-references: EMBL:Z34289; NID:G663007; PIDN:CA84063.1; PID:G663008
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.7%; Score 171.5; DB 2; Length 699;
Best Local Similarity 19.5%; Pred. No. 0.016;
Matches 108; Conservative 84; Mismatches 218; Indels 145; Gaps 19;
Qy 2 DDKGDPNNEAPKAIKPTSKFRKTGWFRRTIAKREGAGADABAD-PLPPPPQOQLGLS 60
Db 136 DBEDQKQVQGVQVQAKAAP-----PKAKSDSDSDSSDEPPKQ-----183
Qy 61 LRSGROPKTERVEQFLTIARRGRSMPSVLSDESGEPTSCPATDAFTASGVSAS 120
Db 184 -----KPKIT-----PVTYKATKAPKPAAPKANGKAASSSS 219
Qy 121 TRGPOSASTAVKRPASSEK-----VKGDDHDTSDSDGLTKELQNR 167
Db 220 SSSSSSSSDSBEKAAATPKTVPKKVQVAKPAVAAATPTTKSSSDSDSDSDEBEQK 279
Qy 168 LRKRRQEP-----TERPLKGIOSRLRKRRREGPAETVGSASDVTGVLPSKQEP 219
Db 280 KPMNKGPSYAPPSAPPKSLGTQPKKAVEKQPVSSSEDSDDSSSBEKKP 339
Qy 220 ENDQGVVQAGKDDRESKLGKAAQDIKDEP-GDLGRPKPECEGYDPNAL-----269
Db 340 FT-KAVVSKATTKPPPAKKAASSSDSDSDSDEDDAPSKPAGTITNSNKPAVTTKSP 398

Qy 270 ----YCIQROPHNNRFMICCDRCBFWHGDVCGISARGLLRNGEDYICPNCTILOVQ 325
Db 399 AVKEAAPKQP-----VGGGQ-----KLTTRKAD-----SSS 425
Qy 326 DETHSETADQOEAK-----WRPGDADGTDCTSI-----GTIEOKSSDQIGKRIEKA 373
Db 426 SEESSSSSEETKQWATYTP-KATAKALSLEPAKQAPQGSRRSDSSSDSSSSEEBEK 484
Qy 374 ANPSGKKLKIQFVIRAPGASKICIGCCCHVAOPDSVYCSNDCILKHAATMFLSGSK 433
Db 485 TSXAVKK-----KPKVAGGAAP-----SKPASA-----KKGKAESSNSSSD 523
Qy 434 BQKPKPKEMKPK-PEKPSLPKCGAQAQIKLSSVHKRPAPKPKKTTVKKAVVVPARSAL 492
Db 524 DSSSEEEKLGKSGSPRQAPANGTSALTQONKAAKNSESEEEKKAAVVKSGSSL 583
Qy 493 CK-----EAACESTTP 503
Db 584 KKRQNEAAKEATP 598

RESULT 8

I51116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I51116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A:Reference number: I51116; MUID:95287814; PMID:7770000
A:Accession: I51116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1110 <JAC>
A:Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632549
C:Superfamily: neurofilament triplet H protein

Query Match 5.7%; Score 171.5; DB 2; Length 1110;
Best Local Similarity 20.0%; Pred. No. 0.025;
Matches 111; Conservative 86; Mismatches 233; Indels 125; Gaps 23;
Qy 2 DDKGDPNNEAPKAIKPTSKFRKTGWFRRTIAKREGAGADABADPLPPPPQOQLGLSL 61
Db 486 DEEEEEEKKEEAEAEEDRG-----RKGEAEAEAEAEAEAEAEAEAEAEAEV 538
Qy 62 RSGROPKTERVEQFLTIARRGRSMPSVLSDESGEPTSCPATDAFTASGVSASSET 121
Db 539 EEAEEETEAEAEAE-----EHEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 587
Qy 122 RSGPOSASTAVKRPASSEKVGKGGDDHDTSDSDGLTKELQNR-----RKK 171
Db 588 -----EAKAEVEEEAEAEAE-----EEBEAEAEAEAEAEAEAEAEAEAEAE 638
Qy 172 EQEPTEPLKGIOSRLRKRRREGPAETVGSASDVTGVLPSKQEPENDQGVVQAGK 231
Db 639 EEEAEAEAEVTSKKATQAEAEVEEEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 691
Qy 232 DDRESKLGKAAQD-----INDEPGLGRPKPECEGYDPNALYCIQROPHNNRPNICC 285
Db 692 EEEEDSKAEADAEDEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 734
Qy 286 DRCEBFWHGDVCGISARGLLRNGEDYICPNCTILOVDETHSETADQOEAKWRPGDA 345
Db 735 -KSEE-----EAAEAKDEAEAEAEAE-----EAEVEEAT-EAEAEKASDDE 777
Qy 346 DGTDTCTSIGTIEOKSS-----DQGIKGRIRKAAANPSGKKLKIQFVIR-----APASKI 398
Db 778 KESE-----EVKSEAPVAPEAKKAPKPAK--PKKAPAKVSEPTSEPEPEKAEVVE 829
Qy 399 GFGCCCHVAQPSVYCSNDCILKHAATMFLSGGKEQPK--PKKMKMKPKPSLPKCG 456

Db 830 KKGKAEAPKPKA-----KPAAKKAEKAPVEKEEPEEPTER---EPKPPAAK-P 876
QY 457 AOAGIKISSVHKPAPEKETTIVKXAVVVPARSE-ALGKEAACESSTPWSADHNNVAVK 515
Db 877 AKAPAK-----PKPAP-KAEAEKZPAPKPAQAKPAPAAEHEDEKEDDEEVEEVVK 930
QY 516 PEKT-----AAPSPS 525
Db 931 PEDAKPVKSKPAPA 944

RESULT 9
T21430
hypoetical protein F26H11.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21430
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21430
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-510 <MIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04195.1; GSPDB:GN00020; CESP:F26H11.3b
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3b
A:Map position: 2
A:Introns: 116/1; 147/3; 288/2; 348/2; 392/1
C:Superfamily: bromodomain homology
P:371-426/Domain: bromodomain homology <BRO>

Query Match 5.7%; Score 170; DB 2; Length 510;
Best Local Similarity 20.4%; Pred. No. 0.014;
Matches 67; Conservative 38; Mismatches 106; Indels 118; Gaps 11;
QY 101 SCPTADTAETASGSEVSASSETSGQPSAST-----AVKRPASSKVKGGDDHDDTSD 153
Db 31 SCPVQDQADASLEISDFAEVLKPFESAEQDWSFGYLLAEQPGTS----- 76
QY 154 SDSGLTLKEL-----QNLRLKREOEPTEPLKIQSLRKKRREGEPAE- 199
Db 77 SDKTTTPIKTIWFKPVPKPGKPRRRRCADREISLAAK---PKAEVKEVINPADI 133
QY 200 TVGSEADPTVEGVLP-----SKQEPENDQGVVSOAGKDDRESKLEGG 241
Db 134 TLGGDTYDVYKEQKTESIATNVSERRRSTANLSKSDDRDKPSQSTAPSKERRTSEP 193
QY 242 AAQIKDEBPQD-----L 254
Db 194 PASHVAFHTPGSATPHDINLSIEHTCOKIPDASKLYIQELCARWTHGCVGVAQTIL 253
QY 255 GRPKPECEG-----YDPNLYCICRQPHNN-RFMICDRCREWFHGDVGVISEARGR 305
Db 254 GLEHWSCEECEIEQERVKDQALYCVCKPYYDKFVYVGDSCQGWTHPECVGTTRAZA- 312
QY 306 LLRNGSDYICPNTTILQVQDTHSETAD 334
Db 313 ---EQAADYNCFAC--REAEVGESEASD 336

RESULT 10
S68191
triadin - human
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 11-Jan-2000
C:Accession: S68191
R:Tsake, N.L.; Eyre, H.J.; O'Brien, R.O.; Sutherland, G.R.; Denborough, M.A.; Foster, P.;
Bur, J. Biochem. 233, 258-265, 1995
A:Title: Molecular cloning of the cDNA encoding human skeletal muscle triadin and its lc
A:Reference number: S68191; MUID:96061957; PMID:7588753
A:Accession: S68191

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-729 <TAS>
A:Cross-references: EMBL:U18985; NID:G882222; PIDN:AAA75315.1; PID:G882223
C:Genetics:
A:Map position: 6q22-6q23
C:Superfamily: histone H1
Query Match 5.7%; Score 169; DB 2; Length 729;
Best Local Similarity 19.3%; Pred. No. 0.023;
Matches 101; Conservative 83; Mismatches 168; Indels 172; Gaps 22;
QY 90 FVSLSDSGEPTSCPATDAETASGSEVSA-----SET-----RSGPOSASTAVKE 134
Db 293 PLPTEQASRPT--PASPALREKEGKKAEKKVTSSTKKKKKEDIKKKSEKETAIDVEKK 350
QY 135 RPASSEKVKGG-----DHDHDSDSGLTLKELQNLRLRRRQ--EPTERP 180
Db 351 EFGKASRTQGVTKIAAQAAKKDKKESKTKKPAEVEQPKGKQKQKGGHVEPAKSP 410
QY 181 LK--GIQSLRKRREGEPAETVGSSEADTVRGVLPSPKQEPENDQGVVSOAGKDDRESKL 238
Db 411 KKEHSPVSDKQVKANTERAKEBIGAVSS---KKAVPKKEKTKTKTVEQIRKEK----- 462
QY 239 EGKAADIDEE-----PGDLCRPKECEGYDPNLYCICRQPHNNRFMICDRCR 289
Db 463 SGTSSILKDEPIKGEKPKVPASLKEKEPETK----- 495
QY 290 EWFHGDVCGISEARGRLRNGEDYICPNTTILQVQDTHSETADQOQAKRPFDAAGTD 349
Db 496 -----KDEKMSKAG---KEVKPKPPQLQGG- 517
QY 350 CTSIGTIEOKSHDQIKGRIEKAANPSGKKLKIPOPIEAPGASKCIGPGGCHVAQPD 409
Db 518 -----KKEKPEPQ-----IKBEKPAISEKVIHKQDIVKP--EKTVSHG-----KPE 558
QY 410 SVYCSNDILKHAATMKFLSSGKQKPKPKKMK-----MKDEKPSLPCG-----A 457
Db 559 -----EKVLQ-----VKAVTIETAKPKTKKAHREPPPSIKTDKPKPTFGTSEVT 608
QY 458 QAGIKTSSVHKPAPKETTIVKXAVVVPARSEALGKEAACESSTPWSAD--HNTNAVK 515
Db 609 ESKGKKTSEISEKESKADMKHLREKVKSTRKESLQNLHNVTKAEKPAFVSKQVEDVPASK 668
QY 516 PEK-----TAAPSPSLLYKCMVHLGVGLDPSRSEFWIAIPW 551
Db 669 KAKEGTEDVSPTKQKSPISPFQCVY-----LDGYNGYGFQPF 706

RESULT 11
I51618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I51618; S57757
R:Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: I51618; MUID:96019267; PMID:7593294
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: mRNA
A:Residues: 1-990 <CAL>
A:Cross-references: EMBL:X88927; NID:G895920; PIDN:CRA61368.1; PID:G895921
C:Genetics:
A:Gene: xNopp180
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.7%; Score 169; DB 2; Length 990;
Best Local Similarity 19.8%; Pred. No. 0.031;
Matches 102; Conservative 54; Mismatches 167; Indels 192; Gaps 16;

QY 7 PSNEAPKAIKPTSKFRKTWGRFTTIAKRG-----AGDAEADPLEPPPPQOQLGLSLR 62
Db 484 PTNSKA-----TPTSKTKTAKPTPTSAKKDDSSSDSSSSSEKKTPT----- 528
QY 63 RSGQPKTERVEQPLTIARRRGSRMPVSLSDSEPTSCPATDAETASEGVSASSTR 122
Db 529 -AKRAAKTTPAKPAKTTAKPAKATTPKPAKATTPKKESSSDSSSDSSSE 587
QY 123 SGPOQASTAVK-----ERPASSKVGDDHDDTSDSDGLTLKELQNLRLRKROE 175
Db 588 K-KSSAKPAKTTKPKATSKPVVASKPVPAKASSSSDSSDSS-----EETT 633
QY 176 PTERLKGIGSLRLKRRREEGPAETVGSSEASTVGVLPKOEPPNDQGVVSQAGKDRE 235
Db 634 KTKPL-----TKLSPAVKTLFPKAESSSDSSSDSSSEKTKP 672
QY 236 SKLEGKAAODIINDEPGDILGRPKCEGYDPNALYICICRQPHNNRMICCDRCCEWFHGD 295
Db 673 AKPPAKSATPVNTKAPAKQKASKASC----- 698
QY 296 CVGISEARGILLERNGEDYICPNCILQVQDETHSTADQQAQKWRPGDADGTDCTSIGT 355
Db 699 -----SDSD----- 702
QY 356 IEQKSSDOGKIRIEKAAANPSGKKLKIQFVIRAPGASKICIGPCCHVAQPDVYCSN 415
Db 703 -----SSSEEG-----KSPKPTGK-----SPAATATAPPK-KNPVAVNKKDPSSSSSD 746
QY 416 DCILKHAAMTKPLSGKQKPKPK-----KKMKPEKSLPKCGAAGIKLSSVHK 468
Db 747 -----SSGDEKQPKQAAMAAKDVQGAQAAPT-PKKAASSSSSDSSDE 791
QY 469 RPAPEKKTIV--KKAVVVPARSEALKEAACESS 501
Db 792 DVSKAKTINAVSKSPVTPKAVPAKSSSESS 826
RESULT 12
JW0057
gravin - human
C:Species: Homo sapiens (man)
C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C:Accession: JW0057
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells
A:Reference number: JW0057; MUID:98269042; PMID:9604001
A:Accession: JW0057
A:Molecule type: mRNA
A:Residues: 1-1684 <SAT>
A:Cross-references: DDBJ:AB003476; NID:G2081606; PIDN:BAAL19927.1; PID:d1020716; PID:g2081606
C:Comment: This protein regulates cell growth.
F:433-439/Region: nuclear location signal
F:522-527/Region: nuclear location signal
F:591-596/Region: nuclear location signal
F:671-676/Region: nuclear location signal
Query Match 5.6%; Score 167.5; DB 2; Length 1684;
Best Local Similarity 19.8%; Pred. No. 0.066;
Matches 136; Conservative 73; Mismatches 218; Indels 261; Gaps 29;
QY 3 DKGDPSNEAP--KAIKPTS-----KSPKTKW-GFRRTT----- 33
Db 169 EGEKQKQKPSKSAESPTFVTSSTGTFKKTPTQCAWAKTKTSFRKPKEDVEASEK 228
QY 34 -----IAKREGAGDAEADPLEPP-----PPQQLG 58
Db 229 KKEQPEKVDTEEDGKAVASEKLTASEQAHPQEPASAEHPRLSAEYKVELFSEQV- 287
QY 59 LSEERSGQ-----PKETERVEQPLTIARRRGSRMPVSLSDSEPTSCPATDA-ET 109
Db 288 -----SGSQGPEBKPAFLATEVFDEKIEVHQEVEVAVHVTVE--ERTEEKTEVEET 340

QY 110 ASEGSVESASSTRSGPOSASTAVKRPASSEKVKVGGDDHDDTSDSDSDGLTLKE----- 163
Db 341 AGSVPAELVEMDAEPQAEPA-KELVKLKETCVSGEDPTQAGADLSPDKVLSKPPSGV 399
QY 164 -----LQNLRLRKROEP-----TERPLKGIQSLRLKRRREG-----PAETVGS 203
Db 400 SEVEMLSQBERMKVQSGPLKLTSTGLKLSKQKQKRGGGDEESGHTQVPAADSPDS 459
QY 204 EASPTVGVLPKQBP-----NDQGVVSQAGKDDRESKLEG----- 240
Db 460 QEEQKSSASSPREPEIITCLEKGLABVQODSEAEAGATSDGKKREGVTPWASPKOMV 519
QY 241 -----KAAODINKDEBPGLGRPKPECEGTDPNALY 270
Db 520 TPCKRVRRPSSDKEDLDKVKASATLSSTESTASEMQEEMKGSVEBKPE----- 569
QY 271 CICEQHNNRMICCDRCCEWFHGDVGVISEARGILLERNGEDYICPNCILQVQDETHS 330
Db 570 -----EPKPK-----VDTSVSWELICVSSSKRA-----RRGSS-----SDEEG 605
QY 331 ETA---DQQAQKWRPGDAD-GTDCSTISGTTIEQKSSBQDQKGRLEKAAANPSGKKLKI 386
Db 606 PKMGGDHOKADEAGKDKETGTDGLAGS--QEHDPGQ--SSSPQAGSPT----- 653
QY 387 PVIEAPGASKICIGPCCHVAQPDVYCSNDCILKHAAMTKPLSGKQKPKPKEMQMK 446
Db 654 -----EGEGVS-----TWESFRLVT-----PRKSKSK 677
QY 447 PEKSLPKCGAAGIKLSSVHKRPAPKKTIVKAVVVPARSEALGK--BAACESSTPS 504
Db 678 LEEKSEDSI-AGSGVEHSTPTDTEPKKESWVSIKKFTPGRRKKRPDQGRQEQAPVEDAGPT 736
QY 505 WASDIN-----YNAVKPEKTA 521
Db 737 GANEDSDVPVAVPLSEYDAVEREKMEA 764
RESULT 13
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <WIL>
A:Cross-references: EXBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CESP:F26H11.3c
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
F:313-368/Domain: bromodomain homology <BRO>
Query Match 5.6%; Score 166; DB 2; Length 452;
Best Local Similarity 21.0%; Pred. No. 0.02;
Matches 55; Conservative 31; Mismatches 82; Indels 94; Gaps 8;
QY 161 LKELQNLRLRKREQFTE-----RPLKGIQSLRLKRRREGPAE-TVGSEAS 206
Db 23 LKKEINKRTRMENAEEQCGLLTPWRKARARPHAAKPAEYKKEVINPADITLGGDTY 82
QY 207 DTVGVLP-----SKQEPNDQGVVSQAGKDDRESKLEGKAAQDIXD 248
Db 83 DYVKEQKPTSIATNVRRRTTSANLSKSDRDKPESOSTAPKSKERTSPASHVAF 142
QY 249 EBPQD-----LGRPKPEC 261

Db 143 HTPGSAHPDINLSIEHCTCOKIFDASKLYIQCBLCARWYHGDGCVGAEQITLGLHWS 202
Qy 262 EG-----YDPNLYCICQPHNN-REMCCDRCREWPHGDGCVGSEARGLLRNGE 312
Db 203 EECIEBERVKDQALYCVCKQFIDDTFYVGCDSQCGWFAPECVGTTRA 258
Qy 313 DYICPNCITLQVQDTHSETAD 334
Db 259 DYNCPACT--REARGYSEASD 278

RESULT 14
B42680
nucleolus-cytoplasm shuttle phosphoprotein - rat
N:Alternate names: Nopp140 protein B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 17-Nov-2000
C:Accession: B42680; A42680; S27889; S30510; S30511
R:Meier, U.; Blobel, G.
Cell 70, 127-138, 1992

A:Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542; PMID:1623516
A:Accession: B42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <ME1>
A:Cross-references: GB:M94288
A:Experimental source: clone pTM6
A:Accession: A42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME3>
A:Cross-references: GB:M94287; NID:9205749; PIDN:AAA41718.1; PID:G205750
R:Meier, U.; Blobel, G.
submitted to the EMBL Data Library, May 1992
A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: S27889
A:Accession: S27890
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:G205751; PIDN:AAA41719.1; PID:G205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME12>
A:Cross-references: EMBL:M94287; NID:G205749; PIDN:AAA41718.1; PID:G205750
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.5%; Score 165.5; DB 2; Length 734;
Best Local Similarity 21.7%; Pred. No. 0.036;
Matches 121; Conservative 75; Mismatches 229; Indels 133; Gaps 25;

Qy 3 DKGPSNEEAPKAIKPTSKFKTWGFRRTTIKRIKREAGDAEADPLPPLPPQQLG 58
Db 117 DSSSDSEEDKAQVPTQK-----AAPAKVASLPQAGKAAKASSSSSSE--- 164
Qy 59 LSLRRSGQPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASEGSVESA 118
Db 165 -SSSEEEBKDKKQVQKAVKQAKAVR--PPPKAESSSESSESSSDEAPQTKPKAA 221
Qy 119 SETRSGPOSASTAVKERPASEKV-----KGGDDHDDTSDSDSGLTTLKELQNLRLRK 171
Db 222 ATAAKAPTKAQTAKAPGPPAKQPKKAANGKAGSSSSSSSSSSSSSD-----S 269
Qy 172 REOSPTERPLKIGISRLKRRKEEPAETVGEASDITVEGVLPSKQEPENDQGVVQAGK 231
Db 270 EEKKAAAPLK-----KTPAKQVQVAKAPVKVTAAPTQKSSSED-----SSSEE 314
Qy 232 DDRSKLEGKAAQDIKDEPGDLGRPKPECEGYDNPALYICRQPHNNRMFCIDRCSEW 291
Db 315 EEEQKPKMKKAGPYSSVPPSPVSLSKSVGAQSPKKA-AACTQPADS-----SADSSSE- 368
Qy 292 FHGDCVGISSEARGLLRNGEDIYICFNCTILQVQDTHSETADQQAQKWRPGDAGTDCT 351

Db 369 -----SDSSSE-----HEKTPAKTV-----SKTPAKPAVKKKARSSSDSD-- 409
Qy 352 SIGTIEKSEDDQIGRIEKAANPSGKKLKFQV-VIEAPGASKCI-----GFGCC 403
Db 410 -----SSED-----EAPAKPVSAKSPKSPKPAVTPKPAKAVATPKQAPAGSG-- 452
Qy 404 HVAQDSVYCSNDCILKHAATAWKFLSSGEQPK-----PKEKMKPKPKPSLP-KCGA 457
Db 453 --QKQSRKAD-----SSSEESSSESEETKSKVTTTPKARVTAK-AAPSLAKQAP 502
Qy 458 QAGIKISSVHKRPAPEKKTETTKKAVVVPARSALG-----KEAACESSTPSWASD 508
Db 503 RAGDSSDSSSESSSEKKTTPKP---PAKKAAGAATVPKPTPVKKAARSSSSSSSSSE 559
Qy 509 HN--YNAVPEKPTAAPSP 524
Db 560 DSSEEEKKPKSKATPKP 577

RESULT 15
H83273
ribonuclease E PA2976 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 31-Dec-2000
C:Accession: H83273
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83273
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1057 <SFO>
A:Cross-references: GB:AE004723; GB:AE004901; NID:99949067; PIDN:AAG06364.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: rne; PA2976
C:Superfamily: ribonuclease E

Query Match 5.5%; Score 165.5; DB 2; Length 1057;
Best Local Similarity 21.5%; Pred. No. 0.053;
Matches 126; Conservative 76; Mismatches 228; Indels 155; Gaps 26;

Qy 7 PSNEEAPKAIKPTSKFKTWGFRRTTIKRIKREAGDAEADPLE--PPPPQQLGLSLERS 64
Db 535 PCHTEAPVEPAKMPPEPSLPQGLVKSLVGLFAGKQPAKPAETSKPAERQTRQDSRN 594
Qy 65 GRQPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASEGSVESASETRSG 124
Db 595 GRQNNRRDGRD---GNRRDEERKPREERAERQPRE-----ERARPRNEESERRRE 644
Qy 125 PQSASTAVKERPASEKVKGGDDHDDTSDSDGLTLKELQNLRLRKREDEPTER---P 180
Db 645 ERA-----ERPAER---QPREGRERAERTFREERQ 675
Qy 181 LKGIQSR-LRKRRKEGPAETVGEASDITVEGVLPSKQEPENDQGVVQAGKQDREK-- 237
Db 676 REGREGEERSEERRER 731
Qy 238 --LEKKA-----AODIKDEEPCDLGRPKPECEGYDNPALYICRQPHNNRMFCIDRC 288
Db 732 AALAEALPNDESLEQDEQDITDGE--RPRRRSGQ-----RRRSNR-----ERQ 775
Qy 289 EEFPHGDCVGISSEARGLLRNGEDIYICFNCTI-----LQVQDE-----THS 330
Db 776 RE-VSGELEG-SEAT-----DNAAPLNTVAAAAAGIAVASEAVANVEQAPATTS 825
Qy 331 ETADQQAQKWRPGDAGTDCTCTGTGIEKSEDDQIGRIEKAANPSGKKLKFQV-VIEAPGASKCI-----GFGCC 381
Db 826 EAASETTA-----SDETASTSEAVETQAGDSEANTGETADIEAFVTVSVVRDEADQST 879

```
QY 382 LKIFQVIEAPGASKCI-----GPGCHVAQPDVYCSNDCILK 420
Db 880 LLVAQATEAPFASVSRESREDAESAVOPATEAAEEVAAPVVEVVAAPSEPAATEEPTPA 939
QY 421 HAA-----ATMKFLSSGKEOKPKPKMKMKPKPSLPKCGAQAQIKISSVHKRPAPKKE 476
Db 940 IAAVPAATGRALNDPREKRRLQREARLAREAAAAAQAQAA---FAVBEIPAVASEE 996
QY 477 TTVKKAUVVPARSEALGKEAACESSTPSWASDHYNNAVKEKTA 521
Db 997 ASAOEE---PAAPOA---EETQADVPSQA-DEAOEAVQAEPEAS 1034
```

Search completed: April 30, 2004, 08:55:59
Job time : 24 secs

XXI

PS	Claim 6; Fig 1C; 27pp; English.	
XX	The invention provides nucleic acids encoding the human and murine death	
CC	inducer-oblierator 1 (DIO-1) polypeptides. The polypeptides can be	
CC	expressed by standard recombinant methodology. The DIO-1 polypeptides,	
CC	agonists and antagonists are used as a medicament for treating diseases	
CC	characterized by an alteration in cell death or by hyperproliferation,	
CC	e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign	
CC	tumours, malignant tumours or hyperproliferative skin disorders. They are	
CC	also useful in the treatment of metabolic, proliferative or inflammatory	
CC	conditions. The present sequence represents the human DIO-1 polypeptide	
XX		
SQ	Sequence 562 AA;	
	Query Match 100.0%; Score 2989; DB 3; Length 562;	
	Best Local Similarity 100.0%; Pred. No. 2.3e-228;	
	Matches 562; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIKREGAGDAADPLEPPPPQOOLGLS 60	
Db	1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIKREGAGDAADPLEPPPPQOOLGLS 60	
QY	61 LRRSGROPKRTVERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASGVSASE 120	
Db	61 LRRSGROPKRTVERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASGVSASE 120	
QY	121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNRLLRRKREQPTERP 180	
Db	121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNRLLRRKREQPTERP 180	
QY	181 LKGIQSLRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240	
Db	181 LKGIQSLRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240	
QY	241 KAAQDIKDEEPCDGLRPKPECEGYDPNALYICICROPHNNRPMICCDRCCEWFHFGDCVGIS 300	
Db	241 KAAQDIKDEEPCDGLRPKPECEGYDPNALYICICROPHNNRPMICCDRCCEWFHFGDCVGIS 300	
QY	301 EARGILLRNGEDYICPNCTIIQVDETHSETADQOEAKWRPDADGTDCTSIGTIEOKS 360	
Db	301 EARGILLRNGEDYICPNCTIIQVDETHSETADQOEAKWRPDADGTDCTSIGTIEOKS 360	
QY	361 SEDQGIKGRIEKAANPSGKKLKIQQPVIEAPGASKICGPGCHVAQPDVSVCSNDCILK 420	
Db	361 SEDQGIKGRIEKAANPSGKKLKIQQPVIEAPGASKICGPGCHVAQPDVSVCSNDCILK 420	
QY	421 HAAATMKFLSSGKEQKPK 480	
Db	421 HAAATMKFLSSGKEQKPK 480	
QY	481 KAVVVPARSEALGKEAACESSTPSWASDHNVAKPEKTAAPSPILLYKMYHLGVGLLD 540	
Db	481 KAVVVPARSEALGKEAACESSTPSWASDHNVAKPEKTAAPSPILLYKMYHLGVGLLD 540	
QY	541 PGRSPWIAIPWACPGIGVAALC 562	
Db	541 PGRSPWIAIPWACPGIGVAALC 562	
RESULT 2		
AA093638	standard; protein; 562 AA.	
XX		
AC	AA093638;	
XX		
DT	26-JUN-2001 (first entry)	
XX		
DE	Human protein sequence SEQ ID NO:13130.	
XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
XX		
OS	Homo sapiens.	
XX		

PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
DF	28-JUL-2000; 2000EP-00116126.	
XX		
PR	29-JUL-1999; 99JP-00248036.	
PR	27-AUG-1999; 99JP-00300253.	
PR	11-JAN-2000; 2000JP-00118776.	
PR	02-MAY-2000; 2000JP-00183767.	
PR	09-JUN-2000; 2000JP-00241899.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
DE	WPI; 2001-318749/34.	
XX		
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-	
PT	length cDNAs defined in the specification, and for the detection and/or	
PT	diagnosis of the abnormality of the proteins encoded by the full-length	
PT	cDNAs.	
XX		
PS	Claim 8; SEQ ID NO 13130; 2537pp + Sequence Listing; English.	
XX		
CC	The present invention describes primer sets for synthesizing 5602 full-	
CC	length cDNAs defined in the specification. Where a primer set comprises:	
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the	
CC	complementary strand of a polynucleotide which comprises one of the 5602	
CC	nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the	
CC	specification. The primer sets can be used in antisense therapy and in	
CC	gene therapy. The primers are useful for synthesizing polynucleotides,	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893	
CC	represent human amino acid sequences; and AAH13629 to AAH13632 represent	
CC	oligonucleotides, all of which are used in the exemplification of the	
CC	present invention	
XX	Sequence 562 AA;	
SQ		
	Query Match 99.7%; Score 2981; DB 4; Length 562;	
	Best Local Similarity 99.8%; Pred. No. 9.8e-228;	
	Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIKREGAGDAADPLEPPPPQOOLGLS 60	
Db	1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIKREGAGDAADPLEPPPPQOOLGLS 60	
QY	61 LRRSGROPKRTVERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASGVSASE 120	
Db	61 LRRSGROPKRTVERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASGVSASE 120	
QY	121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNRLLRRKREQPTERP 180	
Db	121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNRLLRRKREQPTERP 180	
QY	181 LKGIQSLRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240	
Db	181 LKGIQSLRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240	
QY	241 KAAQDIKDEEPCDGLRPKPECEGYDPNALYICICROPHNNRPMICCDRCCEWFHFGDCVGIS 300	

Db 241 KAAQDIKDEEPEDLGRPKPECEGYDPNALYICICROPHNNRPMICCDRCBEWPHGDCVGIS 300
 Qy 301 EARGLLRENGEDYICPNCITLQVQDETHSETADQOEAWRPDADGTDCTSIGTIEOKS 360
 Db 301 EARGLLRENGEDYICPNCITLQVQDETHSETADQOEAWRPDADGTDCTSIGTIEOKS 360
 Qy 361 SEDQIGIKGRIRKAAANPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 420
 Db 361 SEDQIGIKGRIRKAAANPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 420
 Qy 421 HAAATMKFTLSSGKEQPKPKYKWKMPKPSLPKCGAQAQIKISSVHKPEPAPEKKTETVK 480
 Db 421 HAAATMKFTLSSGKEQPKPKYKWKMPKPSLPKCGAQAQIKISSVHKPEPAPEKKTETVK 480
 Qy 481 KAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAPSLLYKCMYHLGVGLLD 540
 Db 481 KAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAPSLLYKCMYHLGVGLLD 540
 Qy 541 PSRSFWIAIPWACPGGLGVAALC 562
 Db 541 PSRSFWIAIPWACPGGLGVAALC 562
 RESULT 3
 AAM80219
 ID AAM80219 standard; protein; 1191 AA.
 AC AAM80219;
 DT 06-NOV-2001 (first entry)
 XX Human protein SEQ ID NO 3865.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX WO200157190-A2.
 XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSB-) HYSBQ INC.
 XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR N-PSDB; AAK53352.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 455-456; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52561), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX SQ Sequence 1191 AA;
 Query Match 81.3%; Score 2431; DB 4; Length 1191;
 Best Local Similarity 92.8%; Pred. No. 1.2e-183;
 Matches 465; Conservative 0; Mismatches 0; Indels 36; Gaps 1;
 Qy 65 GROPKRTERVEQFTIARRGRSRMPVSLSDSGEPTSCPATDAETASBGSVESASSTRSG 124
 Db 22 GROPKRTERVEQFTIARRGRSRMPVSLSDSGEPTSCPATDAETASBGSVESASSTRSG 81
 Qy 125 POSASTAVKRPASSEKVKGGDDHDDTSDSDSGLTTLKELQNLRRKRQEPTEPLKGI 184
 Db 82 POSASTAVKRPASSEKVKGGDDHDDTSDSDSGLTTLKELQNLRRKRQEPTEPLKGI 141
 Qy 185 QSRLEKRREREGPAETVGSASDTVEGVLPSKQEPENDQGVVSOAQKDDRESKLEGKAAQ 244
 Db 142 QSRLEKRREREGPAETVGSASDTVEGVLPSKQEPENDQGVVSOAQKDDRESKLEGKAAQ 201
 Qy 245 DIKDEEPGDLGRPKPECEGYDPNALYICICROPHNNRPMICCDRCBEWPHGDCVGISRARG 304
 Db 202 DIKDEEPGDLGRPKPECEGYDPNALYICICROPHNNRPMICCDRCBEWPHGDCVGISRARG 261
 Qy 305 RLLENGEDYICPNCITLQVQDETHSETADQOEAWRPDADGTDCTSIGTIEOKSSEDO 364
 Db 262 RLLENGEDYICPNCITLQVQDETHSETADQOEAWRPDADGTDCTSIGTIEOKSSEDO 321
 Qy 365 GIKGRIRKAAANPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 388
 Db 322 GIKGRIRKAAANPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 381
 Qy 389 IEAPGASKICIGPGCHVAQPSVYCSNDCILKHAATMKFTLSSGKEQPKPKYKWKMPKE 448
 Db 382 IEAPGASKICIGPGCHVAQPSVYCSNDCILKHAATMKFTLSSGKEQPKPKYKWKMPKE 441
 Qy 449 KPSLPKCGAQAQIKISSVHKRPAPPEKKTETVKAVVVPARSEALGKEAACESSTPSWASD 508
 Db 442 KPSLPKCGAQAQIKISSVHKRPAPPEKKTETVKAVVVPARSEALGKEAACESSTPSWASD 501
 Qy 509 HNYNAVKPEKTAAPSLLYK 529
 Db 502 HNYNAVKPEKTAAPSLLYK 522
 RESULT 4
 AAM79235
 ID AAM79235 standard; protein; 647 AA.
 XX AAM79235;
 XX 06-NOV-2001 (first entry)
 DT Human protein SEQ ID NO 1897.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 OS WO200157190-A2.
 PN 09-AUG-2001.
 PD

XX PF 05-FEB-2001; 2001WO-US004098.
XX PR 03-FEB-2000; 2000US-00496914.
XX PR 27-APR-2000; 2000US-00560875.
XX PR 20-JUN-2000; 2000US-00598075.
XX PR 19-JUL-2000; 2000US-00620325.
XX PR 01-SEP-2000; 2000US-00654936.
XX PR 15-SEP-2000; 2000US-00663561.
XX PR 20-OCT-2000; 2000US-00693325.
XX PR 30-NOV-2000; 2000US-00728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX DR WPI: 2001-476283/51.
XX DR N-PSDB; AAK52368.
XX PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX PT in diagnosis and gene therapy.
XX PS Claim 20; Page 4293-4294; 6221pp; English.
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity slating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
XX CC sequence listing were missing at the time of publication
XX CC Sequence 647 AA;
XX SQ

Query Match 77.3%; Score 2311; DB 4; Length 647;
Best Local Similarity 92.5%; Pred. No. 1.7e-174;
Matches 441; Conservative 0; Mismatches 0; Indels 36; Gaps 1;

QY 89 MPVSLDSGPTSCPATDAETASGVSASGTSRGPQASATVAXRPPASSEKVKGGDDH 148
DB 1 MPVSLDSGPTSCPATDAETASGVSASGTSRGPQASATVAXRPPASSEKVKGGDDH 60
QY 149 DDTSDSDGLTLKELQNLRLRRKREQPTERPLKGIQSLRLKRRREGPAETVGSASPT 208
DB 61 DDTSDSDGLTLKELQNLRLRRKREQPTERPLKGIQSLRLKRRREGPAETVGSASPT 120
QY 209 VEGVLPKQEPNDQGVWVGQKDDRESKLEGAADIDKEEPDGLGRPKPECEGYDPNA 268
DB 121 VEGVLPKQEPNDQGVWVGQKDDRESKLEGAADIDKEEPDGLGRPKPECEGYDPNA 180
QY 269 LYCICQPHNNRPMICDRCEHFWHGDGCVGISSEARGLLERNGEDYICPNCITLQVDET 328
DB 181 LYCICQPHNNRPMICDRCEHFWHGDGCVGISSEARGLLERNGEDYICPNCITLQVDET 240
QY 329 HSTADQOEAKWPRGADGTDCTSIGTIOKSSDQGIKRIEKAANPSGKKKILFQP- 387
DB 241 HSTADQOEAKWPRGADGTDCTSIGTIOKSSDQGIKRIEKAANPSGKKKILFQP- 300
QY 388 -----VIEAPGASKICIGGCCCHVAQPDVY 412
DB 301 PGVPPTQLFVLQVLEIAVSRISAFTHLLHCISKVIEAPGASKICIGGCCCHVAQPDVY 360
QY 413 CSNDCILKHAATAWTKTSLSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 472
DB 361 CSNDCILKHAATAWTKTSLSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 420

QY 473 EKKETTIVKAVVVPARSEALGKEAACESSTPFWASDHNNVAVKDEKTAAPSPLYK 529
DB 421 EKKETTIVKAVVVPARSEALGKEAACESSTPFWASDHNNVAVKDEKTAAPSPLYK 477

RESULT 5
ABG22389
ID ABG22389 standard; protein; 775 AA.
XX AC ABG22389;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #22380.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR N-PSDB; AAS86576.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 52748; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 775 AA;

Query Match 77.1%; Score 2305.5; DB 4; Length 775;
Best Local Similarity 76.8%; Pred. No. 6e-174;
Matches 468; Conservative 22; Mismatches 52; Indels 67; Gaps 13;

QY 1 MDDKGDPNNEAPKAIKTSKEFRWGRFRRTTIAKREGAGDABDLEPPPPQQLGLS 60

QY 537 GLDPSRSFWIAIPWACPL 556
DB 536 GPPGSHHL-----GGCLGL 550
RESULT 7
ID ABG22388 standard; protein; 330 AA.
AC ABG22388;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #22379.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS86575.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 52747; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 330 AA;
Query Match 51.9%; Score 1551; DB 4; Length 330;
Best Local Similarity 89.0%; Pred. No. 1.5e-114;
Matches 290; Conservative 0; Mismatches 0; Indels 36; Gaps 1;
QY 235 ESKLEGKAAQDIKDEEPDGLGRPKPECEGYDPNALYICRQPHNRFMIOCDCREEMFHG 294

DB 1 ESKLEGKAAQDIKDEEPDGLGRPKPECEGYDPNALYICRQPHNRFMIOCDCREEMFHG 60
QY 295 DCVGISEARGHLLRNGEDYICNCTTLOYODSTHSETADOQAKRPGDADGTDCTSIG 354
DB 61 DCVGISEARGHLLRNGEDYICNCTTLOYODSTHSETADOQAKRPGDADGTDCTSIG 120
QY 355 TIEQSSDQGIKRIEKAANPSGKKLKIPOP----- 387
DB 121 TIEQSSDQGIKRIEKAANPSGKKLKIPOP----- 180
QY 388 -----VIEARGASKCIGPGCHVAOPDSVYCSNDCILKHAATMKFLSSGKEOKPK 438
DB 181 TLLHCTISCKVIEAPGASKCIGPGCHVAOPDSVYCSNDCILKHAATMKFLSSGKEOKPK 240
QY 439 PKRNMKKPKPSLPKCGAAGIKISSVHKRPAPEKKEKTTVKKAVVVPARSEALGKEAAC 498
DB 241 PKRNMKKPKPSLPKCGAAGIKISSVHKRPAPEKKEKTTVKKAVVVPARSEALGKEAAC 300
QY 499 ESSTPSWASDHNNYNAVPEKTAAPSP 524
DB 301 ESSTPSWASDHNNYNAVPEKTAAPSP 326
RESULT 8
ID ABG22387 standard; protein; 167 AA.
AC ABG22387;
DT 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #22378.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS86574.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 52746; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AB00010-AB030377 represent novel human diagnostic
 CC amino acid sequences. The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 167 AA;

Query Match 30.0%; Score 898; DB 4; Length 167;
 Best Local Similarity 100.0%; Pred. No. 4e-63;
 Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 304 GLLERNGEDYICPNCTILQVDETHSETADQQAQKWPQDADGDTCTSIGTIQKSSD 363
 DB 1 GLLERNGEDYICPNCTILQVDETHSETADQQAQKWPQDADGDTCTSIGTIQKSSD 60
 QY 364 QGIGRIEKAANPSGKKLKIPOVIEAPGASKICIGPCCCHVAQPDVSVCNDILKHA 423
 DB 61 QGIGRIEKAANPSGKKLKIPOVIEAPGASKICIGPCCCHVAQPDVSVCNDILKHA 120
 QY 424 ATMKFLSSGKKQKPKPKKMKPKPKPSLPKCGAAGIKISSVHKRP 470
 DB 121 ATMKFLSSGKKQKPKPKKMKPKPKPSLPKCGAAGIKISSVHKRP 167

RESULT 9

AAB43724
 ID AAB43724 standard; protein; 181 AA.

AC AAB43724;

DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1169.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
 KW antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 KW vasotropic; antipsoriatic; angiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.

XX Homo sapiens.

XX WO200005350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005882.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.

XX N-PSDB; AAC77933.

XX Novel isolated nucleic acids comprising sequences encoding peptides

XX useful for treating or diagnosing e.g. cancer.

XX Claim 11; Page 1790-1791; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given in

CC AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;
 CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention

XX Sequence 181 AA;

Query Match 29.1%; Score 869; DB 3; Length 181;
 Best Local Similarity 99.4%; Pred. No. 8.8e-61;
 Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 254 LGKPKCEGVDNALYICICQPHNNRFMICDRCRCEWFFHGDGVCISARGELLERNGED 313
 DB 10 LGKPKCEGVDNALYICICQPHNNRFMICDRCRCEWFFHGDGVCISARGELLERNGED 69
 QY 314 YICFNCITLQVDETHSETADQQAQKWPQDADGDTCTSIGTIQKSSDQGIKRIEKA 373
 DB 70 YICFNCITLQVDETHSETADQQAQKWPQDADGDTCTSIGTIQKSSDQGIKRIEKA 129
 QY 374 ANPSGKKLKIPOVIEAPGASKICIGPCCCHVAQ 403
 DB 130 ANPSGKKLKIPOVIEAPGASKICIGPCCCHVAH 164

RESULT 10

ABP05496

ID ABP05496 standard; protein; 108 AA.

XX ABP05496;

XX 24-JUN-2002 (first entry)

XX Human ORFX protein sequence SEQ ID NO:10974.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

XX Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach MD;

Db 1166 QFRKVLQNPSEVLPSTVQS-----ADAEKRLKLGAPRAATSTSEPAVLGVAKKPEGP 1221
 QY 493 GKEAACRSSTPSWASDHYNVAVKP-----BKTA-----APSPSLYKMYHLGVGLDPS 542
 Db 1222 AKLSHPQNTTVQASHQIGISSVRPLAKKDKKTKTPTTVQAPTPN-----RLAAGKPEPV 1274
 QY 543 R 543
 Db 1275 R 1275
 RESULT 12
 AAM78908
 ID AAM78908 standard; protein; 1539 AA.
 XX
 AC AAM78908;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1570.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 05-FEB-2001; 2001WO-US0040398.
 XX
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 (HYSE-) HYSEQ INC.
 XX
 Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZH;
 Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 WPI; 2001-476283/51.
 DR N-PSDB; AAK52041.
 XX
 Nucleic acids encoding polypeptides with cytokine-like activities, useful
 in diagnosis and gene therapy.
 PT
 PT
 PS Claim 20; Page 3896-3899; 6221pp; English.
 XX
 The invention relates to polynucleotides (AAK51456-AAK53435) and the
 encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 cytokine, cell proliferation or cell differentiation or which may induce
 production of other cytokines in other cell populations. The
 polynucleotides and polypeptides are useful in gene therapy, vaccines or
 peptide therapy. The polypeptides have various cytokine-like activities,
 e.g. stem cell growth factor activity, haematopoiesis regulating
 activity, tissue growth factor activity, immunomodulatory activity and
 activin/inhibin activity and may be useful in the diagnosis and/or
 treatment of cancer, leukaemia, nervous system disorders, arthritis and
 inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 (AAK52582) and 3566 (AAM80020) are omitted as the relevant pages from the
 sequence listing were missing at the time of publication
 XX
 Sequence 1539 AA;

Query Match 7.8%; Score 232; DB 4; Length 1539;
 Best Local Similarity 21.8%; Pred. NO. 5.1e-09;
 Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;
 QY 92 SLED-SGEPTSCPATTAETAGSGSVESASETR-SGPQASATAVKERPASSEKVK-GGDH 148
 Db 95 SLSDKSHAHPGCLKEPHHPAQTHGVSHSSQKQCHKPOQAPAMKTNHSHKELEHPGVH 154
 QY 149 DDTSDSDSLTLKE-----LQNLRRKREQPTERPLKGIOSRLKRRRGGPAAVTGSE 204
 Db 155 -----FKBEDKLKKEPEKNLQPRQRSSKSFSLDEPPLFIPDNIAIIRRE-----GS- 202
 QY 205 ASDTVEGVLPKQEPENDQGVVSQAGKDDRESKLEKAAQDIKDEPGDLGRPKDECRGY 264
 Db 203 -----DHSSSFESKY-----MTFSPKQC-GP 222
 QY 265 DPNALYICICROPHNRPMICDRCSEWFGDCVGLSEARGRLRLRNGBDYICPNCITLQV 324
 Db 223 -----CKKPHGNRFVVGCGRCDDNFHGDVGLSLSQAOQMGDEKXYVCVKCC---A 271
 QY 325 QDETHSETAD-----QBEAKWRPGDADGDTCTSIGTIEQKSS-----DQIKGRIEKA 373
 Db 272 EEDKTEILDPDTLENQATVFEHSGDKT-MECKLGLSKHTNDKTYKIDTVKHKVAIL 330
 QY 374 ANPSGKKKLIKIPQVIEAPGASKICIGPGCCCHVAOPDSVYCSNDCILKHAATMKPLSSGK 433
 Db 331 KRESCEGR-----NSSDCRDNEIKKQQLAPLRKMGQPV 363
 QY 434 EOKPKPKERKMKPKPSLPKQGAQAGIKISVHKRPAPERKETTVCXAVVVPARSEALG 493
 Db 364 LPRRSSEKSEKIPKESITVTC---TGEKASKPGTHERQENKKKVKGV-----LN 412
 QY 494 KEAACESSTPS 504
 Db 413 VHPAASAKPS 423
 RESULT 13
 AAM39234
 ID AAM39234 standard; protein; 1728 AA.
 XX
 AC AAM39234;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2379.
 XX
 KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US034263.
 XX
 XX 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58390.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX Example 4; SEQ ID NO 2379; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX

SQ Sequence 1728 AA;

Query Match 7.8%; Score 232; DB 4; Length 1728;
Best Local Similarity 21.8%; Pred. No. 5.9e-09;
Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;
QY 92 SLED-SCEPTSCPATDAETASGVSASSTR-SGPQASSTAVKRPASSEKVK-GGDDH 148
DB 284 SLSKSHAHPCCLKEPHHPAQTGHVSHSQKQCHKPOQAPAMKNTNSHVKEELHPGVGH 343
QY 149 DDTSDSDSGTLKE-----LQNRLLRKRKEQEPTEPLKGIQSLRKRKRREGPAETVGS 204
DB 344 ----FKEEDKLKAKPKENLQPRQRSSKSFSLDEPLFPDNIATIRRE-----GS- 391
QY 205 ASDTVGVLPKQEPENDQGVVSQAGKDRESKLEKGAQDIKDRBPDILGRPKCEGY 264
DB 392 -----DHSSSFESKY-----MWTPSKQC-GF 411
QY 265 DPNALYCIQPHNNRPMI CDRCEWPHGDCVGISEARGLLERMGEDYICPNTILQV 324
DB 412 -----CKPHGNRFWVGCGRDDWFHGDVGLSLSQQQGGEEDKEVYCKCC--A 460
QY 325 QDETHSETAD-----QGEAKWRPDADGTDCTSIGTIEOKSE-----DQIGKRIEKA 373
DB 461 EEDKKTEILDPTLENQATVEFHSQDNT-MECSKGLSLRHTTNDRTKYIDDTVKHKVKIL 519
QY 374 ANPSGKKKLIKFPVIEAPGASKICIGPCCHVAQPSVYCSNDCILKHAATMKFLSGK 433
DB 520 KRESGEGR-----NSSDCRDNEIKKQLAPLRKMGQPV 552
QY 434 EOKPKPKKMKKPEKSLPKCGAAGIKLSSVHKRPAPKPKTTTKYAVVPAREALG 493
DB 553 LPRSSSEKSEKIPKSTTTVTC---TGEKASKPFTGHEKQMKKKKVKRGV-----LN 601
QY 494 KEAACESSTPS 504
DB 602 VHPAASAKPS 612

RESULT 14

AAI79892

ID AAM79892 standard; protein; 1766 AA.

XX

AC AAM79892;

XX 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 3538.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX N-PSDB; AAK53025.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
XX Claim 20; Page 376-377; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 1766 AA;

Query Match 7.8%; Score 232; DB 4; Length 1766;

Best Local Similarity 21.8%; Pred. No. 6.1e-09;

Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;

QY 92 SLED-SCEPTSCPATDAETASGVSASSTR-SGPQASSTAVKRPASSEKVK-GGDDH 148
DB 322 SLSKSHAHPCCLKEPHHPAQTGHVSHSQKQCHKPOQAPAMKNTNSHVKEELHPGVGH 381
QY 149 DDTSDSDSGTLKE-----LQNRLLRKRKEQEPTEPLKGIQSLRKRKRREGPAETVGS 204
DB 382 ----FKEEDKLKAKPKENLQPRQRSSKSFSLDEPLFPDNIATIRRE-----GS- 429
QY 205 ASDTVGVLPKQEPENDQGVVSQAGKDRESKLEKGAQDIKDRBPDILGRPKCEGY 264
DB 430 -----DHSSSFESKY-----MWTPSKQC-GF 449

